

RADIO ENGINEERING and ELECTRONIC PHYSICS

English Edition of

РАДИОТЕХНИКА И ЭЛЕКТРОНИКА

Published by the American Institute of Electrical Engineers
with the aid of a grant from the National Science Foundation

No.2 February 1961

Translated and Produced by Royer and Roger, Inc.

AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS

Established 1884

345 East Forty-Seventh Street
New York 17, N. Y.

Warren H. Chase, President
N.S. Hibshman, Executive Secretary
C.E. Dean, Technical Vice President, Communications
W.F. Denkhau, Director of Publications
L.G. Abraham, Chairman, Communications Division

The English edition of RADIO ENGINEERING AND ELECTRONIC PHYSICS is published by the American Institute of Electrical Engineers with the aid of a grant from the National Science Foundation. © 1961 by American Institute of Electrical Engineers. Also published under the same arrangement are the Russian electronic journals RADIO ENGINEERING and TELECOMMUNICATIONS.

RADIO ENGINEERING AND ELECTRONIC PHYSICS

(РАДИОТЕХНИКА И ЭЛЕКТРОНИКА)

Publication of the Institute of Radio Engineering and Electronic Physics,
Academy of Sciences of the USSR

Translated and Produced
by
Royer and Roger, Inc.



Translation Editor: Herbert Dern, Columbia University

AIEE REVIEW COMMITTEE FOR RADIO ENGINEERING AND ELECTRONIC PHYSICS

Leonard S. Schwartz
New York University College of Engineering
Chairman

| | | | |
|-----------------|-------------------|------------------|------------------|
| A.W. Bickley | A. Burr Fontaine | W. Miller | C.A. Stutt |
| W.P. Birkemeier | F.E. Froelich | Harry Rowe Mimno | G.C. Sziklai |
| T.T.W. Bucher | Paul H. Gleichauf | W.W. Peterson | Joseph Vogelmann |
| J.L. Callahan | G.S. Glinski | B. Reiffen | G.M. White |
| G.R. Cooper | Bernard Harris | W.G. Schmidt | F.B. Wood |
| W.A. Depp | R.K. Hellmann | Herbert Sherman | H.L. Yudkin |
| R.G. Enticknay | D.E. Higginbotham | D.L. Solomon | |

Subscriptions to Radio Engineering and Electronic Physics should be sent to AIEE
Special Subscription Department
41 East 28th Street, New York 16, New York

1961 Subscription rates:

| | \$ | £ |
|---------------------------------------|-------|----|
| Individuals | 28.50 | 10 |
| Libraries, institutes, govt. agencies | 57.00 | 20 |

12 issues per annum comprising approximately 1900 pages

NOISE STORAGE IN DELAYED FEEDBACK DEVICES

Yu. S. Lezin

The noise power at the output of delayed-feedback storage devices is determined by the "slowly varying magnitudes" method.

INTRODUCTION

The most important element of optimal filters for sequences of pulsed signals [1] is the summer with a feedback circuit having a delay equal to the repetition period of these signals. Equality of the repetition period and the delay time in the feedback circuit can be attained with sufficient accuracy for practical purposes by use of the methods employed in systems for moving-target indication [2]. However, this kind of summing device is self-exciting because its positive feedback coefficient is unity.

A practical approximation to such a device is a memory with positive feedback and a delay time T . For stable operation the memory's feedback coefficient should be less than unity. Figure 1a gives a block diagram of the memory, consisting of a summer; an attenuator with transfer coefficient m (equal to the feedback coefficient at sufficiently low frequencies), a delay line with delay time T and a filter F with transfer function

$$K^*(\omega) = \frac{1}{1 + j\omega\delta} = \frac{1}{1 + j \frac{\omega}{2\pi\Delta F}},$$

where δ is the time constant of the filter and ΔF is its bandwidth at the 3 db point. This transfer function coincides (to within a constant factor) with the frequency characteristic of an uncompensated resistance coupled amplifier in the high-frequency region.

All the elements of the memory besides the filter have uniform amplitude-frequency characteristics extending at least up to the frequency at which the modulus of the filter's transfer function becomes negligibly small. Let us assume that all the elements are linear.

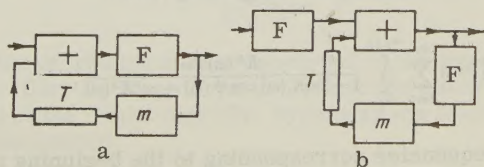


Figure 1. Block diagram of a delayed feedback memory.

A memory for pulsed signals in such delayed-feedback devices has been thoroughly analyzed in Ref. 3.

In the present article we consider a memory for conventional white noise with a power spectrum

$$F_1(\omega) = a.$$

The results obtained are also accurate for the passage of conventional white noise through a filter F and subsequent storage in a delayed feedback device (Fig. 1b).

1. NOISE STORAGE IN A SINGLE MEMORY WITH DELAYED FEEDBACK

The transfer function of such a memory is given by

$$K_1^*(\omega) = \frac{K^*(\omega)}{1 - mK^*(\omega)e^{-j\omega T}} = \frac{K(\omega)e^{-j\arctg \omega \delta}}{1 - mK(\omega)e^{-j\psi(\omega)}},$$

and the square of its modulus

$$K_1^2(\omega) = \frac{K^2(\omega)}{1 - 2mK(\omega)\cos\psi(\omega) + m^2K^2(\omega)},$$

where

$$K(\omega) = \frac{1}{\sqrt{1 + \omega^2\delta^2}} \quad (1)$$

is the modulus of the transfer function of the filter F;

$$\psi(\omega) = \omega T + \arctg \omega \delta. \quad (2)$$

Therefore, the stored noise in the memory with delayed feedback has, at the output, the power

$$\sigma_2^2 = \int_0^\infty F_1(\omega) K_1^2(\omega) d\omega = a \int_0^\infty \frac{K^2(\omega) d\omega}{1 - 2mK(\omega)\cos\psi(\omega) + m^2K^2(\omega)}. \quad (3)$$

Evaluation of this integral is ordinarily quite difficult. Therefore, we make an approximation, using the circumstance, peculiar to the devices we are considering, that the time constant of the filter is very small in comparison with the delay time:

$$\delta \ll T. \quad (4)$$

Thanks to this we can apply the method of slowly varying magnitudes [4] to the evaluation of Eq. (3). It is apparent that the second term in the denominator of the integrand contains a rapidly oscillating component $\cos\psi(\omega) = \cos(\omega T + \arctg \omega \delta)$, but the remaining terms of the integrand, in comparison, change slowly with frequency. On a change of frequency equivalent to one period of the oscillating component, the remaining terms in the denominator and numerator are practically unchanged. We break up the integration in Eq. (3) into parts, each of which corresponds to one period of the oscillating component:

$$\sigma_2^2 = a \sum_{l=0}^{\infty} \int_{\omega_l}^{\omega_{l+1}} \frac{K^2(\omega) d\omega}{1 - 2mK(\omega)\cos\psi(\omega) + m^2K^2(\omega)},$$

where ω_l and ω_{l+1} are the frequencies corresponding to the beginning and end of the l -th period.

$$\psi(\omega_l) = 2\pi l, \quad \psi(\omega_{l+1}) = 2\pi(l+1).$$

As a consequence of inequality (4) the value of this period is approximately equal to the repetition frequency.

$$\Omega_l = \omega_{l+1} - \omega_l \simeq \Omega = \frac{2\pi}{T}.$$

We make a change of variable under the integral sign and let $x = \omega - \omega_l$.

$$\sigma_2^2 = a \sum_{l=0}^{\infty} \int_0^{\Omega_l} \frac{K^2(\omega_l + x) dx}{1 - 2mK(\omega_l + x) \cos \psi(\omega_l + x) + m^2 K^2(\omega_l + x)}.$$

As a consequence of inequality (4), for $0 < x < \Omega_l$ we have

$$K(\omega_l + x) \simeq K(\omega_l), \\ \cos \psi(\omega_l + x) \simeq \cos xT$$

and therefore,

$$\begin{aligned} \sigma_2^2 &\simeq a \sum_{l=0}^{\infty} \int_0^{\Omega_l} \frac{K^2(\omega_l) dx}{1 - 2mK(\omega_l) \cos xT + m^2 K^2(\omega_l)} \simeq \\ &\simeq \frac{2a}{T} \sum_{l=0}^{\infty} K^2(\omega_l) \int_0^{\pi} \frac{dy}{1 - 2mK(\omega_l) \cos y + m^2 K^2(\omega_l)}. \end{aligned}$$

Using Eq. (3.451.2) of Ref. 5 we obtain

$$\sigma_2^2 \simeq a\Omega \sum_{l=0}^{\infty} \frac{K^2(\omega_l)}{1 - m^2 K^2(\omega_l)} \simeq a \sum_{l=0}^{\infty} \frac{K^2(\omega_l)}{1 - m^2 K^2(\omega_l)} \Delta\omega_l.$$

This result can also be obtained by the stationary phase method [6].

Since the magnitude of the l -th period of the function $\cos \psi(\omega)$ in view of the inequality (4) is extremely small in comparison to the width of the passband, and consequently, also, in comparison to the width of the noise spectrum at the output, then the last sum can be approximated by the integral:

$$\sigma_2^2 \simeq \int_0^{\infty} a \frac{K^2(\omega)}{1 - m^2 K^2(\omega)} d\omega. \quad (5)$$

The integrand $a \frac{K^2(\omega)}{1 - m^2 K^2(\omega)}$ can be considered as a power spectrum of the stored noise,

averaged over a period of the oscillating factor $\cos(\omega T + \arctg \omega \delta)$. Substituting Equation (1) into Equation (5) and evaluating, we obtain

$$\sigma_2^2 \simeq \frac{\pi^2 a \Delta F}{\sqrt{1 - m^2}}. \quad (6)$$

Thus, the power of the stored noise is proportional to the intensity of the noise at the input and passband of the filter F . In order to determine the power of the unstored noise, that is the noise arriving at the output directly, bypassing the feedback circuit, it is sufficient to set $m = 0$ in the last expression:

$$\sigma_{20}^2 = \pi^2 a \Delta F.$$

Therefore, the coefficient of noise storage in the delayed feedback memory, in terms of power, is equal to

$$g_1 \simeq \frac{1}{\sqrt{1-m^2}}.$$

If a rectangular approximation is made to the amplitude-frequency characteristic of the delayed feedback loop of the memory, i.e., if we consider the filter to be an ideal low-pass filter, then the noise storage coefficient is seen to be equal to [7].

$$g' = \frac{1}{1-m^2}.$$

The dependence of the noise storage coefficient on the magnitude of the feedback coefficient, both in the case considered and in an ideal filter, are presented in Curves 1 and 2, respectively, of Fig. 2. Noise in the real memory investigated is accumulated to a much smaller degree than in an ideal memory with a feedback loop having a rectangular amplitude-frequency characteristic. This is explained by the narrowing of the power spectrum of the noise when it is circulated in the feedback loop of a real memory. In an idealized memory the noise power spectrum is uniformly attenuated only by a factor of m each time it is circulated, and its width remains unchanged.

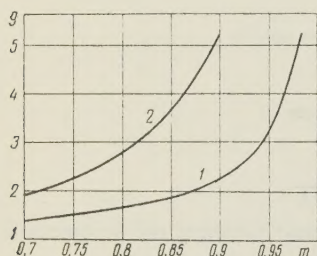


Figure 2. Noise storage coefficient (1) for a single memory (2) with an ideal filter, as functions of the feedback coefficient.

2. NOISE STORAGE IN TWO AND THREE SERIES-CONNECTED MEMORIES WITH DELAYED FEEDBACK

The noise at the output of two series-connected identical memories, each of which is encompassed by a feedback with a delay time T , and arranged as in the block diagram presented in Fig. 1 has the power

$$\mathcal{J}_3^2 = \int_0^\infty F_1(\omega) K_1^4(\omega) d\omega = a \int_0^\infty \frac{K^4(\omega) d\omega}{[1 - 2mK(\omega) \cos \psi(\omega) + m^2 K^2(\omega)]^2}.$$

For an approximate evaluation of this integral we average the comb power spectrum of the output noise over a period of the rapidly oscillating factor $\cos \psi(\omega)$. As a result of this we obtain

$$\mathcal{J}_3^2 \simeq a \int_0^\infty \frac{[1 + m^2 K^2(\omega)] K^4(\omega)}{[1 - m^2 K^2(\omega)]^3} d\omega.$$

Substituting Eq. (1) into the resulting expression, transposing, and applying Eq. (3.191.3) of Ref. 5 we determine the power of the stored noise

$$\mathcal{J}_3^2 \simeq \frac{a\pi^2 \Delta F}{4} \frac{2 + m^2}{(1 - m^2)^{3/2}}, \quad (7)$$

and then also the noise storage coefficient of two memories with delayed feedback,

$$g_2 = \frac{\sigma_3^2}{\sigma_3^2(m=0)} \simeq \frac{2+m^2}{2(1-m^2)^{1/2}}. \quad (8)$$

This is less than the noise storage coefficient of two memories with rectangular amplitude-frequency characteristics in the delayed feedback loop:

$$g'' = \frac{1+m^2}{(1-m^2)^3}.$$

The dependence of the noise storage coefficients of two memories on the feedback coefficient for the case considered and for rectangular amplitude frequency characteristics are presented in Fig. 3 (Curves 1 and 2, respectively). The ratio of the noise storage coefficients for one and two memories with delayed feedback is equal to

$$\lambda_2 = \frac{g_2}{g_1} = \frac{2+m^2}{2(1-m^2)^2}.$$

The dependence of this ratio on the feedback coefficient is given by Curve 3, Fig. 3. This ratio greatly exceeds the noise storage coefficient for a single memory. The explanation of this is that on passing through the first memory the noise becomes noticeably correlated, and, because of this, storage in the second memory behaves differently and thus, reaches a much greater level.

In a similar manner we calculate the noise power at the output of three identical series-connected memories with delayed feedback.

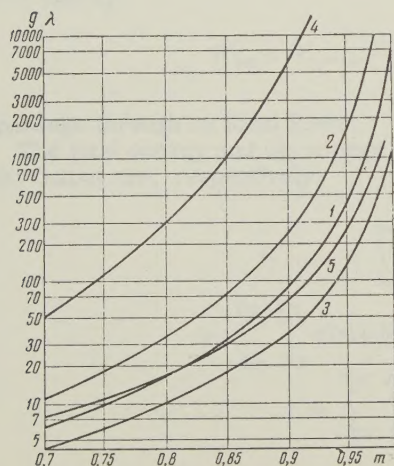
$$\sigma_4^2 \simeq \frac{3a\pi^2\Delta F}{64} \frac{8+24m^2+3m^4}{(1-m^2)^{1/2}} \quad (9)$$

and the noise storage coefficient for the three memories

$$g_3 = \frac{\sigma_4^2}{\sigma_4^2(m=0)} \simeq \frac{8+24m^2+3m^4}{8(1-m^2)^{1/2}}. \quad (10)$$

The dependence of the latter on the feedback coefficient is shown in Curve 4, Fig. 3. The ratio of the noise-storage coefficients for two and three memories with delayed feedback is equal to

$$\lambda_3 = \frac{g_3}{g_2} \simeq \frac{8+24m^2+3m^4}{4(2+m^2)(1-m^2)^2}$$



and is shown as a function of the feedback coefficient in Curve 5, Fig. 3.

From a consideration of the curves given, it is possible to conclude that the ratio of the noise storage coefficients for two and three memories sharply increase as the feedback coefficient

Fig. 3. Noise storage coefficients for two and three memories and the ratio of these coefficients as functions of the feedback coefficient.

approaches unity and even at $m = 0.9$ are 291 and 142,600, respectively, which is a consequence of the strong correlation of the noise of the two

preceding memories with delayed feedback. This correlation increases as the feedback coefficient approaches unity.

CONCLUSION

The noise power at the output of one, two and three memories with delayed feedback and frequency characteristics in the feedback loop coinciding with those of an uncompensated resistance coupled amplifier in the high-frequency region have been calculated by the "slowly varying magnitudes" method. Applying the method described we may make calculations for delayed feedback memories utilizing filters having a different frequency characteristic, and, also, for systems involving frequency filters and delayed feedback memories.

The calculations show that noise is stored in the delayed feedback memories considered to a lesser level than with ideal filters with rectangular amplitude-frequency characteristics. This is explained by the fact that the power spectrum of the noise contracts each time it is circulated in the feedback loop, when the filter possesses non-uniform amplitude-frequency characteristics.

Noise stored in the second, and especially, in the third memory behaves differently than in the first, and reaches an extremely high level, due to the strong correlation it acquires on passage through the first and second delayed feedback memories.

REFERENCES

1. Yu. S. Lezin, Optimal Filters for Sequences of Pulsed Signals, NDVSh (Radio Engineering and Electronics) 1958, 1, 3, 20.
2. P. A. Bakulev, Radar Methods for Moving-Target Indication, published by Oborongiz, 1958.
3. L. A. Morugin, Storage of Pulsed Signals in Delayed Feedback Devices, Radio Engineering and Electronics, 1960, 5, 12, 1885.
4. I.S. Gonorovskii, Fundamentals of Radio Engineering, Svyazizdat, 1957, p. 498.
5. I.M. Ryzhik, I.S. Gradshtein, Tables of Integrals, Sums of Series and Products, GITTL, 1951[
6. S. Goldman, Frequency Analysis, Modulations and Noise, translated from English under the editorship of G.S. Gorelik, Foreign Literature Press, 1951, p. 115.
7. Communication Theory, edited by W. Jackson, Butterworth Scientific Publications. London, 1953, p. 216-230.

Original article submitted to the editors

28 March 1960

Re-submitted after amendment on

9 September 1960

LINEAR CORRECTION OF TRANSIENTS IN SYSTEMS WITH FINITE PASSBAND

P. Ya. Nudel'man

The report discusses the problem of decreasing the transient phenomena caused by the passage of a pulse through a linear system with a finite passband.

Let us examine an ideal filter

$$Z(\omega) = \begin{cases} 1, & |\omega| \leq \Omega, \\ 0, & |\omega| > \Omega. \end{cases} \quad (1)$$

We shall assume that phase shift is zero, which does not diminish the general applicability of the discussion.

If a pulse $\varphi(t)$ is applied to the input of such a filter, on the basis of the Duhamel formula we obtain at the output

$$f(t) = \int_{-T/2}^{T/2} \frac{1}{\pi} \frac{\sin \Omega(t-\tau)}{t-\tau} \varphi(\tau) d\tau. \quad (2)$$

Here $(-T/2, T/2)$ is the duration interval of pulse $\varphi(t)$. All values of $f(t)$ corresponding to instants of time outside the interval $(-T/2, T/2)$ will be known as after-effect:

Let us now determine the form of the impulse $\varphi(t)$ of T duration having a minimal after-effect energy:

$$W_{ae} = W_{-\infty, \infty} - W_{-T/2, T/2} = \int_{-\infty}^{\infty} f^2(t) dt - \int_{-T/2}^{T/2} f^2(t) dt \quad (3)$$

in passage through an ideal filter.

The total energy and the energy in the interval $(-T/2, T/2)$ of the function $f(t)$ at the filter output are, respectively,

$$\begin{aligned} W_{-\infty, \infty} &= \int_{-\infty}^{\infty} \left[\int_{-T/2}^{T/2} \frac{1}{\pi} \frac{\sin \Omega(t-\tau)}{t-\tau} \varphi(\tau) d\tau \right]^2 dt = \\ &= \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \varphi(\tau) \varphi(s) \int_{-\infty}^{\infty} \frac{1}{\pi^2} \frac{\sin \Omega(t-\tau)}{t-\tau} \frac{\sin \Omega(t-s)}{t-s} dt d\tau ds = \\ &= \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \varphi(\tau) \varphi(s) \frac{1}{\pi} \frac{\sin \Omega(\tau-s)}{\tau-s} d\tau ds, \end{aligned} \quad (4)$$

$$W_{-T/2, T/2} = \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \varphi(\tau) \varphi(s) \frac{1}{\pi^2} \int_{-T/2}^{T/2} \frac{\sin \Omega(t-\tau)}{t-\tau} \frac{\sin \Omega(t-s)}{t-s} dt d\tau ds. \quad (5)$$

In order to find the pulse yielding a minimum in Eq. (3) let us investigate the functional

$$U[\varphi] = \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \varphi(\tau) \varphi(s) \frac{1}{\pi} \frac{\sin \Omega(\tau-s)}{\tau-s} d\tau ds - \\ - \lambda \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \varphi(\tau) \varphi(s) \frac{1}{\pi^2} \int_{-T/2}^{T/2} \frac{\sin \Omega(t-\tau)}{t-\tau} \frac{\sin \Omega(t-s)}{t-s} dt d\tau ds. \quad (6)$$

It is necessary to find the function (henceforth designated as $\bar{\varphi}(t)$ for which (see, for example, Ref. 1) $\delta U[\varphi]=0$ with the smallest value of the parameter $\lambda = \lambda_{\min}$.

Introducing the arbitrary function $\zeta(t)$ and the parameter ϵ , we obtain the following equation for the determination of λ_{\min} and $\bar{\varphi}(t)$:

$$\frac{\partial}{\partial \epsilon} U[\bar{\varphi}(t) + \epsilon \zeta(t)]_{\epsilon=0} = 0. \quad (7)$$

Equation (7) converts to the form

$$\int_{-T/2}^{T/2} \bar{\varphi}(\tau) \frac{1}{\pi} \frac{\sin \Omega(t-\tau)}{t-\tau} d\tau - \\ - \lambda \int_{-T/2}^{T/2} \bar{\varphi}(\tau) \frac{1}{\pi^2} \int_{-T/2}^{T/2} \frac{\sin \Omega(t-s)}{t-s} \frac{\sin \Omega(s-\tau)}{s-\tau} ds d\tau = 0. \quad (8)$$

Equation (8) is satisfied only by the eigenfunctions of a homogeneous integral equation of the Fredholm type with symmetrical positively defined kernel:

$$\bar{\varphi}(t) = \lambda \int_{-T/2}^{T/2} \frac{1}{\pi} \frac{\sin \Omega(t-\tau)}{t-\tau} \bar{\varphi}(\tau) d\tau \quad (9)$$

with values of λ coinciding with the characteristic numbers of Eq. (9).

This equation was derived by M.S. Gurevich [2] for the pulse carrying the maximum part of the total energy in the given bandwidth.

Thus, the pulse with the maximum concentrated spectrum (in the energy sense) during passage through an ideal filter provides minimum after-effect (also in the energy sense).

Let us note one property of $\bar{\varphi}(t)$, flowing from the physical sense of Eq. (9), which may be regarded as the Duhamel formula [compare with Eq. (2)]. In passing through the system of Eq. (1) the pulse $\bar{\varphi}(t)$ does not change its shape in the interval $(-T/2, T/2)$ but only decreases its amplitude by λ_{\min} times and is analytically extended from the interval $(-T/2, T/2)$ over the entire time axis $(-\infty, \infty)$.

If, instead of an ideal filter, we examine another narrow band system $K(\omega)$ with the same cutoff frequency Ω then it is evident that the pulse $\bar{\varphi}(t)$ must be replaced by pulse $\psi(t)$ so that

$$\Psi(\omega) K(\omega) = \bar{\Phi}(\omega), \quad (10)$$

where

$$\bar{\Phi}(\omega) = \int_{-T/2}^{T/2} \bar{\varphi}(t) e^{-i\omega t} dt; \quad (11)$$

$$V(\omega) = \int_{-T/2}^{T/2} \psi(t) e^{-i\omega t} dt. \quad (12)$$

In other words, the pulse must be "matched" with the system so that there is obtained at its output a spectrum, corresponding to the function $f(t)$, with the best energy concentration in time.

This may be assumed as the basis of another approach to the problem. Let it be necessary to find $F(\omega)$ — the spectrum (at the output of a system with finite bandwidth) equal to zero at $|\omega| > \Omega$ — such that the energy of the after-effect

$$W_{ae} = \int_{-\infty}^{\infty} f^2(t) dt = \int_{-T/2}^{T/2} f^2(t) dt \quad (13)$$

is minimum; that is, we arrive at the problem discussed in Ref. 2, the only difference being that the time and frequency axes are interchanged. Here

$$f(t) = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) e^{i\omega t} d\omega. \quad (14)$$

Instead of a determination of $\bar{\varphi}(t)$, we seek its Fourier transform $\bar{\Phi}(\omega)$ in the frequency interval $(-\Omega, \Omega)$, which obviously is a unique definition of $\bar{\Phi}(\omega)$ over the entire frequency axis and, consequently, of $\bar{\varphi}(t)$.

Investigation of the stability of the functional

$$\int_{-\infty}^{\infty} \left[\frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) e^{i\omega t} d\omega \right]^2 dt - \lambda \int_{-T/2}^{T/2} \left[\frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) e^{i\omega t} d\omega \right]^2 dt \quad (15)$$

or, which is the same,

$$\frac{1}{2\pi} \int_{-\Omega}^{\Omega} |F(\omega)|^2 d\omega - \frac{\lambda}{2\pi^2} \int_{-\Omega}^{\Omega} \int_{-\Omega}^{\Omega} \frac{\sin \frac{T}{2}(\omega - y)}{\omega - y} F(\omega) F(y) d\omega dy \quad (16)$$

leads to the equation

$$\bar{\Phi}(\omega) = \lambda \int_{-\Omega}^{\Omega} \frac{1}{\pi} \frac{\sin \frac{T}{2}(\omega - y)}{\omega - y} \bar{\Phi}(y) dy. \quad (17)$$

As stated above, we have designated by $\bar{\Phi}(\omega)$ that one of the admissible functions $F(\omega)$ in which the functional (15) assumes a stable value. Comparing Eqs. (9) and (17), we arrive at the equality

$$\bar{\varphi}\left(\frac{r}{\Omega}\right) = \alpha \bar{\Phi}\left(\frac{2r}{T}\right), \quad -\frac{T\Omega}{2} \leq r \leq \frac{T\Omega}{2}, \quad (18)$$

which is at least valid with the simple characteristic number λ_{\min} ; α may be determined by means of the Rayleigh theorem.

In Ref. 2 M.S. Gurevich performs an approximate solution of Eq. (9). According to Ref. 2, the shape of $\bar{\varphi}(t)$ with small values of $\Omega T/2$ is almost rectangular and with an increase in $\Omega T/2$ there is observed a considerable decrease in $\bar{\varphi}(t)$ toward the extremities of the interval $(-T/2, T/2)$ and with $\frac{\Omega T}{2} > \frac{3}{2} \pi$ on the plot of $\bar{\varphi}(t)$ there appear at the

extremities of the interval $(-T/2, T/2)$ oscillations about the zero level. For values of $\Omega T/2$ from the interval $(\pi, \frac{3}{2}\pi)$, $\bar{\varphi}(t)$ is close to the function $\cos \frac{\pi}{T} t$.

In summation, it may be said that the pulse $\bar{\varphi}(t)$ is a maximally frequency-compressed pulse. In passage through an ideal filter it is converted into a maximally time-compressed function repeating the shape of $\bar{\varphi}(t)$ in the interval of the initial duration. The spectrum of the pulse, $\bar{\Phi}(\omega)$, when λ_{\min} is simple, is associated with the pulse by Eq. (18), that is, is its dual in the sense of the Fourier transform. A system with cutoff frequency Ω yields at the output a maximally time-concentrated function if the product of the spectrum of the pulse and the system transfer function coincides in the passband with $\bar{\Phi}(\omega)$.

Hence it is evident that compensation must reduce to selection of such a pulse with the given linear system (or of a compensating filter with a given code signal) that the spectrum of the signal at the output coincides with $\bar{\Phi}(\omega)$ within the passband of the system.

Let us mention again that we have discussed only idealized linear systems with infinite attenuation in the stopband. Energy was chosen as the criterion in evaluating transient intensity.

REFERENCES

1. R. Courant, D. Hilbert, *Methods of Mathematical Physics*, 1, GITTL, 1951, p. 181.
2. M.S. Gurevich, *Radio Engineering and Electronics*, 1956, 1956, 1, 3, 313.

Submitted to the editors 7 June 1960

CALCULATION OF CHANNEL CAPACITY IN TWO-PATH COMMUNICATION

Lin Hai-Ts'uan

The article discusses the capacity of a two-path communications channel with randomly varying parameters (propagation amplitude and propagation time) in the presence of Gaussian additive noise. A formula is derived for the capacity of the channel in the case of normally distributed random amplitudes and a uniform noise spectrum.

INTRODUCTION

Much attention is currently being devoted to the investigation of the theory of communications channels with randomly varying parameters. These investigations include Refs. 1 and 2 and others.

Reference 2 investigates the capacity of a two-path channel (frequently encountered in practice) and derives a simple formula for determining the capacity in the case of a uniform noise spectrum. Instead of a real two-path communications channel, the present paper examines an idealized model of a channel in which the amplitudes of the two waves are considered to be nonfluctuating.

CALCULATION OF CAPACITY

Let $\xi(t)$ be the transmitted signal at the input of a channel having finite average power P_c and finite bandwidth $(0, W)$, within the limits of which the propagation conditions for any of the spectral components of $\xi(t)$ may be considered identical. In addition, it is considered that two-path propagation of radio waves occurs.

Under these conditions the signal $\eta(t)$ at the channel output may be represented as follows:

$$\eta(t) = a_1(t) \xi[t - \tau_1(t)] + a_2(t) \xi[t - \tau_2(t)] + \zeta(t). \quad (1)$$

Here $a_1(t)$, $a_2(t)$, $\tau_1(t)$ and $\tau_2(t)$ are random functions, independent of $\xi(t)$ and $\zeta(t)$, which describe amplitude and arrival time fluctuations of the corresponding waves; $\zeta(t)$ is additive noise, independent of $\xi(t)$.

As shown in Ref. 2 the calculation of the capacity of the two-path channel described by Eq. (1) in the general case is a complex nonlinear problem,

The physical nature of the propagation of medium and short waves permits substantial simplification of Eq. (1). It may be replaced by an idealized model of the channel

$$\eta(t) = a_1 \xi(t - \tau_1) + a_2 \xi(t - \tau_2) + \zeta(t), \quad (2)$$

where a_1 and a_2 are fixed, nonrandom numbers equal to the root mean square values of the processes $a_1(t)$ and $a_2(t)$; τ_1 and τ_2 are random quantities whose distributions coincide with the one dimensional distributions of the processes $\tau_1(t)$ and $\tau_2(t)$. In Ref. 2 there is derived a formula for determining the capacity of just such a channel. Below we examine a model which may be described by the relationship

$$\eta(t) = x \xi(t - \tau_1) + y \xi(t - \tau_2) + \zeta(t). \quad (3)$$

Here x and y are random quantities whose distributions coincide with the one dimensional distributions of the processes $a_1(t)$ and $a_2(t)$. Such a model provides a closer approximation to the real communications channel.

Let us assume that the transmitted signal $\xi(t)$ is a random stationary process. Let the additive noise $\zeta(t)$ be a Gaussian random stationary process with zero mean. The mean noise power will be designated by P_n .

The channel capacity of the model of Eq. (3) may be represented by the expression [3]

$$C'' = \max_{\{\xi(t)\}} M \bar{I}(\xi, \eta/x, y, \tau_1, \tau_2), \quad (4)$$

where $\bar{I}(\xi, \eta/x, y, \tau_1, \tau_2)$ is the information rate; averaging is performed over all x , y , τ_1 , τ_2 . If τ_1 and τ_2 are considered to be independent of x and y , then

$$C'' = \max_{\{\xi(t)\}} \int \int P(x, y) dx dy \int \int P(\tau_1, \tau_2) \bar{I}(\xi, \eta/x, y, \tau_1, \tau_2) d\tau_1 d\tau_2. \quad (5)$$

As shown in Ref. 2 in the presence of additive noise with a uniform spectrum the capacity of the model of Eq. (2)

$$C'' = \max_{\{\xi(t)\}} M\bar{I}(\xi, \eta / \tau_1, \tau_2) = \\ = F \log \frac{1}{2} \left[1 + (a_1^2 + a_2^2) \frac{P_c}{P_n} + \sqrt{1 + (a_1^2 - a_2^2)^2 \frac{P_c^2}{P_n^2} + 2(a_1^2 + a_2^2) \frac{P_c}{P_n}} \right]. \quad (6)$$

Hence in the case of the model of Equation (3)

$$C'' = \iint P(x, y) F \log \frac{1}{2} \left[1 + (x^2 + y^2) \frac{P_c}{P_n} + \sqrt{1 + (x^2 - y^2)^2 \frac{P_c^2}{P_n^2} + 2(x^2 + y^2) \frac{P_c}{P_n}} \right] dx dy = \\ = \frac{F}{\ln 2} \iint P(x, y) \Phi(x, y, P_c, P_n) dx dy, \quad (7)$$

where

$$\Phi(x, y, P_c, P_n) = \ln \frac{1}{2} \left[1 + (x^2 + y^2) \frac{P_c}{P_n} + \sqrt{1 + (x^2 - y^2)^2 \frac{P_c^2}{P_n^2} + 2(x^2 + y^2) \frac{P_c}{P_n}} \right].$$

Let us assume that x and y are independent random quantities with normal distribution. In this case

$$C'' = \frac{F}{\ln 2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_1}} e^{-\frac{(x-a_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi\sigma_2}} e^{-\frac{(y-a_2)^2}{2\sigma_2^2}} \Phi(x, y, P_c, P_n) dx dy, \quad (8)$$

where a_1, a_2, σ_1 and σ_2 are, respectively, means and standard deviations of the random quantities x and y .

If the dispersion of the random amplitudes is small, then, expanding $\Phi(x, y, P_c, P_n)$ in a Taylor series in terms of $x = a_1, y = a_2$, we obtain

$$\Phi(x, y, P_c, P_n) = \Phi(a_1, a_2, P_c, P_n) + (x - a_1) \frac{\partial}{\partial x} \Phi(a_1, a_2, P_c, P_n) + \\ + (y - a_2) \frac{\partial}{\partial y} \Phi(a_1, a_2, P_c, P_n) + \frac{1}{2!} \left\{ (x - a_1)^2 \frac{\partial^2}{\partial x^2} \Phi(a_1, a_2, P_c, P_n) + \right. \\ + 2(x - a_1)(y - a_2) \frac{\partial^2}{\partial x \partial y} \Phi(a_1, a_2, P_c, P_n) + \\ \left. + (y - a_2)^2 \frac{\partial^2}{\partial y^2} \Phi(a_1, a_2, P_c, P_n) \right\} + \dots \quad (9)$$

Substituting Eq. (9) in Eq. (8) and integrating, we obtain

$$C'' = \frac{F}{\ln 2} \left\{ \Phi(a_1, a_2, P_c, P_n) + \frac{\sigma_1^2}{2!} \frac{\partial^2}{\partial x^2} \Phi(a_1, a_2, P_c, P_n) + \right. \\ \left. + \frac{\sigma_2^2}{2!} \frac{\partial^2}{\partial y^2} \Phi(a_1, a_2, P_c, P_n) + \dots \right\}. \quad (10)$$

The resulting series is an approximate expression for the capacity of the two-path channel with randomly varying parameters in the presence of additive noise with a uniform spectrum. Equation (10) is applicable in the medium- and short-wave ranges.

Let us examine a few particular cases.

1. In the case where fluctuation is absent (i.e., $\sigma_1 = \sigma_2 = 0$) the expression for C'' is transformed into the expression for C' obtained in Ref. 2.

2. When the ratio P_c/P_n is small

$$C'' \simeq \frac{F}{\ln 2} \frac{P_c}{P_n} (a_1^2 + a_2^2 + \sigma_1^2 + \sigma_2^2). \quad (11)$$

This means that for small signal-noise power ratios the capacity of a two-path channel has the same value as a one-path channel with a transfer constant (for power) $a_1^2 + a_2^2 + \sigma_1^2 + \sigma_2^2$.

3. If the ratio P_c/P_n is large,

$$C'' \simeq \begin{cases} F \log a_1^2 \frac{P_c}{P_n} & \text{with } a_1^2 \geq a_2^2, \\ F \log a_2^2 \frac{P_c}{P_n} & \text{with } a_1^2 \leq a_2^2. \end{cases} \quad (12)$$

The result coincides with that obtained in Ref. 2 for this case. This is evident since, with a large ratio P_c/P_n and small dispersions σ_1 and σ_2 , the amplitude fluctuations are negligible.

4. In the case where the second wave is absent (i.e., $a_2 = \sigma_2 = 0$) and the output signal is represented in the form $\eta(t) = x\xi(t - \tau_1) + \zeta(t)$, we have

$$C'' \simeq \frac{F}{\ln 2} \frac{P_c}{P_n} (a_1^2 + \sigma_1^2). \quad (13)$$

In conclusion I wish to thank V.I. Siforov and B.S. Tsybakov for their constant assistance in this work.

REFERENCES

1. V.I. Siforov, Toward a theory of radio communications channels with multipath propagation, Transactions of the A.S. Popov Scientific and Technical Society for Radio Engineering and Telecommunications, No II, 1958.
2. B.S. Tsybakov, Concerning the capacity of two-path communications channels, Radio Engineering and Electronics, 1958, 4, 7, 1116.
3. R.L. Dobrushin, Transmission of information over a channel with feedback, Theory of Probability and Its Applications, 1958, 3, 4, 395.

Submitted to the editors 14 July 1960

CAPACITY OF ONE-PATH AND TWO-PATH COMMUNICATIONS CHANNELS WITH POLARIZATION FADING

Lin Hai-Ts'uan

The report discusses the capacity of one-path and two-path communications channels with polarization fading. Formulas are derived for determining the capacity of such channels in the presence of Gaussian noise with uniform spectral density. It is shown that the decrease in capacity of a one-path channel due to polarization fading does not exceed 14.5%.

INTRODUCTION

The capacity of two-path communications channels with interference fading was discussed in Refs. 1 and 2. However, in short-wave radio communications, in addition to interference fading, there exists polarization fading of the received signal, which also affects the capacity of short-wave communications channels. Hence it is of interest to calculate the capacity of communications channels with polarization fading.

1. ONE-PATH COMMUNICATIONS CHANNEL

Let us first examine a one-path communications channel. As is known, under the influence of the earth's magnetic field a plane-polarized wave reaching the ionosphere is divided into two elliptically polarized waves. Along the propagation path of the radio waves, the direction of the electric field vector of each wave changes continually due to fluctuation variations in the concentration of electrons. The change in direction of the plane of polarization will lead to fading of the signals.

Let it be assumed that reception is achieved with a horizontal dipole and that the direction of the electric field vector of the incoming wave relative to the dipole assumes various equiprobable values. Mathematically this is described by the following:

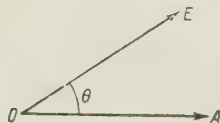


Fig. 1. Relation between directions of the electric field vector E of the incoming wave and the horizontal dipole A .

$$W(\theta) = \begin{cases} \frac{1}{2\pi} & \text{for } 0 < \theta < 2\pi, \\ 0 & \text{for } \theta < 0, \theta > 2\pi, \end{cases} \quad (1)$$

where θ is the angle between direction of the electric field vector of the incoming wave and the dipole Fig. (1); $W(\theta)$ is the probability density of the angle θ .

Let $\xi(t)$ be the transmitted signal with mean power P_c and $\eta(t)$, the received signal at the

channel output. Then

$$\eta(t) = x\xi(t - \tau) + \zeta(t), \quad (2)$$

where $\xi(t)$ is the additive noise, the mean power of which is designated as P_n ; x is a random quantity characterizing the component of the electric field vector in the dipole direction.

If a_1 represents the transfer constant (for electric field strength), then in the given case $x = a_1 \cos \theta$. As is shown, if the angle θ has a uniform distribution, then the probability density of the random quantity x has the form

$$P(x) = \frac{1}{\pi \sqrt{a_1^2 - x^2}} \quad \text{for } |x| < 1. \quad (3)$$

For a channel operating as described in Eq. (2) the capacity

$$C'_n = \int P(x) F \log \left(1 + x^2 \frac{P_c}{P_n} \right) dx. \quad (4)$$

Substituting Eq. (3) in Eq. (4), we obtain

$$C'_n = \int_{-a_1}^{a_1} \frac{1}{\pi \sqrt{a_1^2 - x^2}} F \log \left(1 + x^2 \frac{P_c}{P_n} \right) dx. \quad (5)$$

After integrating, we have

$$C'_n = F \log \frac{4}{2} \left(1 + \frac{a_1^2 P_c}{2 P_n} + \sqrt{1 + a_1^2 \frac{P_c}{P_n}} \right). \quad (6)$$

This equation permits the calculation of the capacity of a one-path channel with polarization fading in the case of Gaussian noise with a uniform spectrum. It follows from this that the capacity depends on the bandwidth occupied by the signal and the ratio of the received-signal power to the noise power $a_1^2 P_c / 2 P_n$ at the channel output.

Let us examine a few particular cases.

1. When the signal power is small in comparison with the noise power the capacity of the channel discussed is

$$C'_n \simeq F \log \left(1 + \frac{a_1^2 P_c}{2 P_n} \right). \quad (7)$$

Hence it is seen that the capacity of the channel with polarization-fading coincides in this case with the capacity of a one-path channel with a transfer constant (for power) $a_1^2 / 2$.

2. When the ratio P_c / P_n is large

$$C'_n \simeq F \log a_1^2 \frac{P_c}{P_n}. \quad (8)$$

It follows from this equation that the capacity of the channel discussed coincides with the corresponding quantity for a one-path channel having a power transfer constant a_1^2 .

Now let us compare the capacity C'_n in the presence of polarization fading with the capacity C_1 of a one-path channel in the absence of polarization fading. The average power of the received signal at the channel output (designated as P'_c) is assumed to be identical in both cases.

From Shannon's formula

$$C_1 = F \log \left(1 + \frac{P'_c}{P_n} \right). \quad (9)$$

where $P'_c = P_c x^2$ (see Ref. 1). In the given case the quantity $\overline{x^2}$, equal to $a_1^2/2$, is easily found from its probability density. Hence

$$C_1 = F \log \left(1 + \frac{a_1^2 P_c}{2 P_n} \right). \quad (10)$$

Figure 2 shows the dependence of ratio C'_n/C_1 on P'_c/P_n , where $P'_c = a_1^2 P_c/2$. It is easily seen that with a change in P'_c/P_n from 0 to ∞ , the ratio C'_n/C_1 decreases from 1 to 0.855 and then rises and approaches unity. Hence it follows that the capacity of a one-path channel decreases due to polarization fading by not more than 14.5% at any signal-noise ratio.

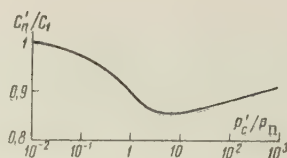


Fig. 2. Capacity C'_n/C_1 as a function of signal-noise ratio (P'_c/P_n)

2. TWO-PATH COMMUNICATIONS CHANNEL

The preceding section discussed the capacity of a one-path channel with polarization fading. If, due to single and double reflection from the ionosphere, there arrive at the receiving point two waves in each of which polarization fading occurs, then there is formed a two-path channel with polarization fading. Let us investigate the capacity of such a channel.

As shown in Ref. 2, the capacity of a two-path channel with randomly varying amplitude and phase in the presence of white noise may be calculated by means of the equation

$$C''_n = \iint P(x, y) C'(x, y) dx dy,$$

where

$$C'(x, y) = F \log \frac{1}{2} \left[1 + (x^2 + y^2) \frac{P_c}{P_n} + \sqrt{1 + (x^2 - y^2)^2 \frac{P_c^2}{P_n^2} + 2(x^2 + y^2) \frac{P_c}{P_n}} \right].$$

We shall assume that the quantities x and y are independent. Substituting the expressions for their probability densities in the expression for C''_n , we obtain

$$C''_n = \int_{-a_2}^{a_2} \int_{-a_1}^{a_1} \frac{1}{\pi \sqrt{a_1^2 - x^2}} \frac{1}{\pi \sqrt{a_2^2 - y^2}} F \log \frac{1}{2} \left[1 + (x^2 + y^2) \frac{P_c}{P_n} + \sqrt{1 + (x^2 - y^2)^2 \frac{P_c^2}{P_n^2} + 2(x^2 + y^2) \frac{P_c}{P_n}} \right] dx dy. \quad (11)$$

The equation permits the calculation of the capacity of a two-path channel in which magneto-ionic division occurs.

Let us examine a few particular cases of Eq. (11).

1. When the ratio P_c/P_n is small

$$C''_n \approx \frac{F}{\ln 2} \frac{P_c}{P_n} \frac{1}{2} (a_1^2 + a_2^2). \quad (12)$$

The capacity of the channel discussed coincides with the corresponding quantity for the channel investigated by Shannon with a power transfer constant of $\frac{1}{2} (a_1^2 + a_2^2)$.

2. In the same case where $a_1/a_2 \gg 1$ and $P_c/P_n \ll 1$,

$$C_n'' \simeq F \log \frac{1}{2} \left(1 + \frac{a_1^2 P_c}{2 P_n} + \sqrt{1 + a_1^2 \frac{P_c}{P_n}} \right) + \frac{F}{\ln 2} \frac{P_c}{P_n} \frac{a_2^2}{2}. \quad (13)$$

In the limiting case with $a_2 \rightarrow 0$, the quantity C_n'' goes over to C_n' (Eq. (6)), which was to be expected.

3. When the ratio P_c/P_n is large

$$C_n'' \simeq \begin{cases} F \log a_1^2 \frac{P_c}{P_n} & \text{for } a_1^2 > a_2^2, \\ F \log a_2^2 \frac{P_c}{P_n} & \text{for } a_1^2 < a_2^2. \end{cases} \quad (14)$$

From Eq. (14) and the corresponding equations of Refs. 1 and 2 it follows for this case that with a large signal-noise ratio the capacity does not depend on whether or not the signal amplitudes fluctuate.

In conclusion I wish to express my thanks to V.I. Siforov and B.S. Tsybakov for their constant assistance and attention in this work.

REFERENCES

1. B.S. Tsybakov, Capacity of two-path communications channels, *Radio Engineering and Electronics*, 1959, 4, 7, 1116.
2. Lin' Khay-tsyuan', Calculation of channel capacity in two-path communications channels, *Radio Engineering and Electronics*, 1961, 6, 2, 197.

Submitted to the editors 14 July 1960

THE PERTURBATION METHOD AND ITS RELATION TO THE EXACT DIFFRACTION METHOD IN SOLVING PROBLEMS OF INCOHERENT SCATTERING

A. V. Shabel'nikov

The exact solution of the problem of the diffraction of a plane electromagnetic wave at a dielectric sphere is expanded into a power series in the small parameter $\alpha = \Delta n k_0 a$. It is established that the solution obtained by the perturbation method is identical to the first term of the expansion. The limits of applicability of the perturbation method are evaluated. The perturbation method is used in solving the problem of the scattering of a plane electromagnetic wave at a dielectric sphere with constant and variable ϵ . The solution with variable ϵ under certain conditions, imposed in the form of a function $\epsilon(r)$, is identical with the solution obtained in the scattering of radio waves at statistical discontinuities of the troposphere.

INTRODUCTION

In solving the problem of incoherent scattering of radio waves at discontinuities of the troposphere the perturbation method [1-4] is usually used. It is distinguished by simplicity and permits the calculation of the scatter field from volumes of any shape with any dependence of the dielectric constant on position. However this method is approximate and hence may be used only under definite conditions.

It is the purpose of the present paper to determine the relation between the perturbation method and the exact diffraction method and to establish the limits of applicability of the perturbation method. For this purpose the exact solution of the problem of diffraction of a plane electromagnetic wave at a sphere is expanded into a series in terms of powers of $\alpha = \Delta n k_0 a$ and it is established that the solution obtained by the perturbation method is identical with the first term of the expansion.

In the conclusion the perturbation method is used in solving the problem of the scattering of a plane electromagnetic wave at a dielectric sphere with constant and variable ϵ . The solution with constant ϵ is identical with the exact solution of the diffraction of a plane electromagnetic wave at a sphere obtained by Jobst [5] on the same assumptions under which the perturbation method is applicable.

The solution with variable $\epsilon(r)$, where r is the distance from the coordinate origin to the reradiating point of the sphere with an appropriate selection of the function $\epsilon(r)$ is identical with the solution obtained in the scattering of radio waves at statistical discontinuities of the troposphere by the use of the correlation function $R(r)$.

1. RELATION OF THE PERTURBATION METHOD TO THE EXACT DIFFRACTION METHOD

Let us examine the propagation of a plane electromagnetic wave in a medium in which the dielectric constant is a function of position. In this case the equation satisfied by the

electric vector \vec{E} will be [2]

$$\Delta \vec{E} + k_0^2 \vec{E} = + \frac{4\pi}{c} \vec{J} i k_0, \quad (1)$$

where

$$\epsilon = \epsilon_2 + S \Delta \epsilon; \quad S \ll 1; \quad k_0 = \frac{2\pi}{\lambda}; \quad \epsilon_2 = \text{const}; \quad (2)$$

\vec{J} is the current density in the antenna; $\Delta \epsilon$ in the general case is a function of position. Substituting Eq. (2) into Eq. (1), we obtain the discontinuous wave equation

$$\Delta \vec{E} + k_0^2 \epsilon_2 \vec{E} = - \Delta \epsilon S k_0^2 \vec{E} + \frac{4\pi}{c} \vec{J} i k_0, \quad (3)$$

the solution of which may be written in the following form:

$$\vec{E}(\vec{r}_0) = \vec{E}_0^*(\vec{r}_0) - S \frac{1}{4\pi\epsilon_2} \int_V \Delta \epsilon k_0^2 \vec{E}(\vec{r}) \frac{e^{ik_0 R}}{R} dV, \quad (4)$$

where \vec{r}_0 is the vector coordinate of the point of observation; $\vec{E}_0^*(\vec{r}_0)$ is the field created by the transmitting antenna at the point of observation in the presence of a homogeneous atmosphere; V is the reradiating volume; R is the distance between the point of observation and a variable point of reradiation in the volume V (see Figure); k_0 is the wave number in a homogeneous medium.

In solving the integral equation let us use the method of successive approximations [6], for which it is necessary to find the solution of Eq. (4) in the form of a series in terms of positive integral powers of S :

$$\vec{E}(\vec{r}_0) = \vec{E}_0(\vec{r}_0) + S \vec{E}_1(\vec{r}_0) + S^2 \vec{E}_2(\vec{r}_0) + \dots + S^p \vec{E}_p(\vec{r}_0) + \dots \quad (5)$$

Substituting the series of Eq. (5) into Eq. (4), integrating term by term and equating coefficients of like powers of S on both sides of the resulting equality, we obtain the formulas for the sequential determination of $\vec{E}_p(\vec{r}_0)$:

$$\vec{E}_0(\vec{r}_0) = \vec{E}_0^*(\vec{r}_0), \quad (6)$$

$$\vec{E}_1(\vec{r}_0) = \frac{1}{4\pi\epsilon_2} \int_V \Delta \epsilon k_0^2 \vec{E}_0(\vec{r}) \frac{e^{ik_0 R}}{R} dV, \quad (7)$$

$$\vec{E}_2(\vec{r}_0) = \frac{1}{4\pi\epsilon_2} \int_V \Delta \epsilon k_0^2 \vec{E}_1(\vec{r}) \frac{e^{ik_0 R}}{R} dV, \quad (8)$$

$$\dots \dots \dots \vec{E}_p = \frac{1}{4\pi\epsilon_2} \int_V \Delta \epsilon k_0^2 \vec{E}_{p-1}(\vec{r}) \frac{e^{ik_0 R}}{R} dV, \quad (9)$$

where

$$R = \sqrt{r^2 + r_0^2 - 2rr_0 \cos \xi};$$

$\cos \xi = \cos \Theta \cos \Theta_0 + \sin \Theta \sin \Theta_0 \cdot \cos(\varphi - \varphi_0)$; r , Θ , φ and r_0 , Θ_0 , φ_0 are the coordinates of the points of reradiation and the point of observation in the spherical coordinate system.

We will note that Eq. (7) is the classical expression for the scatter field, which is usually derived in all theories of scattering in using the perturbation method. In order to establish the relation between the perturbation method and the exact diffraction method it is necessary to precisely evaluate the integral of Eq. (7), wherein in this case $\Delta \epsilon = \text{const}$.

Usually in scatter theory in evaluating integrals of this form the quantity R is expanded into a series, thereby taking advantage of the fact that the calculation of the field is being made in the wave zone. Hence in evaluating the perturbation method we must know the value of the integral in the vicinity of the scatter volume, that is, at points where r_0 is of the order of or even less than r .

In this case the expansion of R into a series has no meaning and it is hence necessary to find a method permitting precise evaluation of an integral of the form of Eq. (7).

Thus, it is necessary to evaluate the integral

$$\vec{E}_1 = -\frac{\Delta \epsilon k_0^2}{4\pi} \vec{E}_0 \int_V e^{ik_1 r} \cos \Theta \frac{e^{ik_1 R}}{R} dV, \quad (10)$$

where \vec{E}_0 is the amplitude of the plane wave. For this purpose let us expand the integrand into a series in terms of Bessel functions and Legendre polynomials [7]

$$\begin{aligned} \vec{E}_1 = & -\frac{\Delta \epsilon k_0^2}{4\pi} \vec{E}_0 \int_V \left\{ \sum_{n=0}^{\infty} i^n (2n+1) \Psi_n(k_2 r) P_n(\cos \Theta) \right\} \times \\ & \times \left\{ ik_2 \sum_{n=0}^{\infty} (2n+1) \Psi_n(k_2 r) h_n^{(1)}(k_2 r_0) P_n(\cos \xi) \right\} dV, \end{aligned} \quad (11)$$

where

$$\begin{aligned} \Psi_n(k_2 r) &= \sqrt{\frac{\pi}{2k_2 r}} J_{n+\frac{1}{2}}(k_2 r); \\ h_n^{(1)}(k_2 r_0) &= \sqrt{\frac{\pi}{2k_2 r_0}} H_{n+\frac{1}{2}}^{(1)}(k_2 r_0), \end{aligned}$$

while the function $P_n(\cos \xi)$, in accordance with the addition theorem [8], may be written in the form of a series

$$\begin{aligned} P_n(\cos \xi) &= P_n(\cos \Theta) P_n(\cos \Theta_0) + \\ &+ 2 \sum_{m=1}^n \frac{(n-m)!}{(n+m)!} P_n^m(\cos \Theta) P_n^m(\cos \Theta_0) \cos m(\varphi - \varphi_0). \end{aligned} \quad (12)$$

Substituting Eq. (12) into Eq. (11) and performing the successive integrations for φ , Θ and r , we obtain the final expression for the radial component of the scatter field:

$$\begin{aligned} E_{1r} = & \frac{\pi \Delta n k_0 a^2}{2r_0} E_0 \sum_{n=0}^{\infty} i^n (2n+1) h_n^{(1)}(k_2 r_0) P_n^1(\cos \Theta_0) \cos \varphi_0 \times \\ & \times \left\{ J_{n+\frac{1}{2}}^2(k_2 a) - J_{n-\frac{1}{2}}(k_2 a) J_{n+\frac{3}{2}}(k_2 a) \right\}, \end{aligned} \quad (13)$$

where E_0 is the amplitude of the wave incident at the dielectric volume V ; a is the radius of the dielectric sphere; k_2 is the wave number in the medium.

Let us now examine the exact solution of the diffraction of a plane electromagnetic wave at a dielectric sphere of radius a . Let us represent the radial component of the scatter field outside the sphere as follows [10]:

$$E_{1r} = -\frac{E_0}{k_2 r_0} \sum_{n=0}^{\infty} i^{n+1} (2n+1) b_n^r h_n^{(1)}(k_2 r_0) P_n^1(\cos \Theta_0) \cos \varphi_0, \quad (14)$$

where

$$b_n^r = -\frac{\Psi_n(y) [x \Psi_n(x)]' - \frac{x^2}{y^2} \Psi_n(x) [y \Psi_n(y)]'}{h_n^{(1)}(y) [x \Psi_n(x)]' - \frac{x^2}{y^2} \Psi_n(x) [y h_n^{(1)}(y)]'}; \quad (15)$$

$$x = N\rho; \quad y = \rho; \quad N = \frac{x}{y}; \quad \rho = k_2 a;$$

the primes indicate differentiation with respect to the argument; $x - y = \alpha = \Delta n k_0 a$; $\Delta n = n_1 - n_2$; n_2 is the refractive index of the medium; n_1 is the refractive index of the sphere.

Let us designate

$$A_n(x) = \frac{[x \Psi_n'(x)]'}{x^2 \Psi_n'(x)}.$$

whence Eq. (15) is rewritten as follows:

$$b_n^r = \beta_n(x) = \frac{[y \Psi_n'(y)]' - A_n(x) y^2 \Psi_n'(y)}{[y h_n^{(1)}(y)]' - A_n(x) y^2 h_n^{(1)}(y)}. \quad (16)$$

Since $x = y + \alpha$, the function $\beta_n(y + \alpha)$ may be expanded into a Taylor series:

$$\beta_n(y + \alpha) = \beta_n(y) + \frac{\alpha}{1!} \beta_n'(y) + \dots + \frac{\alpha^q}{q!} \beta_n^{(q)}(y + \gamma \alpha), \quad (17)$$

where γ is a number lying between 0 and 1.

Calculations show that

$$\begin{aligned} \beta_n(y) &= 0, \\ \beta_n'(y) &= -i \{ y^2 [\Psi_n'(y) \Psi_n''(y) - (\Psi_n''(y))^2] - 2 [\Psi_n'(y)]^2 - y \Psi_n'(y) \Psi_n'(y) \}. \end{aligned} \quad (18)$$

After appropriate transpositions, Eq. (14) in the first approximation takes the form

$$\begin{aligned} E_{1r} &= \frac{\pi \Delta n k_0 a^2}{2r_0} E_0 \sum_{n=0}^{\infty} i^n (2n+1) h_n^{(1)}(k_2 r_0) P_n^1(\cos \Theta_0) \cos \varphi_0 \times \\ &\times \left\{ J_{n+\frac{1}{2}}^2(k_2 a) - J_{n-\frac{1}{2}}(k_2 a) J_{n+\frac{3}{2}}(k_2 a) \right\}. \end{aligned} \quad (19)$$

Comparison of Eqs. (13) and (19) reveals their complete identity, which leads to the following important conclusion. The expression for the scatter field usually used in the theory of noncoherent scattering represents the first term of the expansion of the exact solution of the diffraction problem into a series of powers of $\alpha = \Delta n k_0 a$. Evaluation of the integrals (8), (9) etc. by the above method has permitted establishing the following identities:

$$E_{2r} = -\frac{E_0}{k_2 r_0} \sum_{n=0}^{\infty} i^{n+1} (2n+1) \frac{\alpha^2}{2!} \beta_n''(y) h_n^{(1)}(k_2 r_0) P_n^1(\cos \Theta_0) \cos \varphi_0, \quad (20)$$

$$E_{pr} = -\frac{E_0}{k_2 r_0} \sum_{n=0}^{\infty} i^{n+1} (2n+1) \frac{\alpha^p}{p!} \beta_n^{(p)}(y) h_n^{(1)}(k_2 r_0) P_n^1(\cos \Theta_0) \cos \varphi_0. \quad (21)$$

Equations (19), (20) and (21) show that the series (5) is an exact solution of the problem posed, wherein the rate of convergence of this series is determined by the rate of convergence of series (17). Series (17) permits the determination of the limits of applicability of the perturbation method. For this purpose it is necessary to satisfy the following inequalities:

$$\frac{\alpha^2}{2} \beta_n''(y + \gamma \alpha) \ll \alpha \beta_n'(y) \quad (22)$$

or

$$\frac{\alpha}{2} \frac{\beta_n''(y + \gamma\alpha)}{\beta_n'(y)} \ll 1. \quad (23)$$

Evaluation of the quantity $A_1 = |\beta_n''(y + \gamma\alpha)/\beta_n'(y)|$ shows that it does not exceed unity for any values of the argument. This indicates that inequality (23) may be rewritten as follows:

$$\frac{\Delta n k_0 a}{2} \ll 1. \quad (24)$$

Inequality (24) defines the limits of applicability of the perturbation method. Chernov [9] also arrives at such an inequality in applying the perturbation method to an examination of the scattering of sound waves at fluctuations in the density and velocity of the medium.

Let us determine the number of terms of series (5) which must be taken in order to obtain the solution of integral equation (4) with an error not exceeding a given value of μ_n . It will be determined from the following equality:

$$\frac{\alpha^p}{p!} \beta_n^{(p)}(y + \gamma\alpha) = \mu_n. \quad (25)$$

The procedure for establishing the limits of applicability of the perturbation method may easily be extended to dielectric volumes of any shape. For this purpose let us examine an arbitrary volume V' with a dielectric constant differing by $\Delta\epsilon$ from the surrounding medium. Let us inscribe around this volume a sphere with radius a , where $2a$ is the distance between the two outermost points of the volume V' . The sphere volume V is

$$V = V' + V'', \quad (26)$$

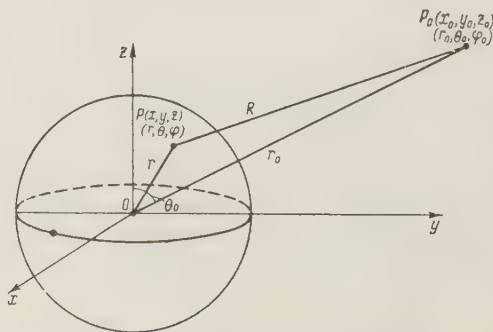
where V'' is the volume required to bring V' up to the volume of the sphere. From Eq. (26) it follows that

$$\frac{1}{4\pi\epsilon_2} \int_{V'} \Delta\epsilon k_2^2 \vec{E}_0(\vec{r}) \frac{e^{ik_2 R}}{R} dV \ll \vec{E}_1(\vec{r}_0), \quad (27)$$

that is, condition (24) is also applicable for the volume V' .

2. APPLICATION OF THE PERTURBATION METHOD IN SOLVING THE PROBLEM OF SCATTERING OF A PLANE ELECTROMAGNETIC WAVE AT A DIELECTRIC SPHERE WITH CONSTANT AND VARIABLE ϵ

A. Let us demonstrate the advantage of the perturbation method over the exact diffraction method in the problem of scattering of a plane electromagnetic wave by a dielectric sphere with constant ϵ . The coordinate systems and geometric relations used in this case are shown in the figure.



Geometric relations in the scattering of radio waves at a dielectric sphere.

It was shown in Section 1 that in the first approximation the field scattered by the sphere may be represented as follows:

$$\vec{E}_1 = \frac{\Delta \varepsilon k_0^2}{4\pi} \int_V \vec{E}^0(\vec{r}) \frac{e^{ik_z R}}{R} dV, \quad (28)$$

where $\vec{E}^0(\vec{r})$ is the incident plane wave propagated in the positive z direction:

$$\vec{E}^0(\vec{r}) = \vec{E}_0 e^{ik_z z}. \quad (29)$$

Substituting Eq. (29) into Eq. (28), the field scattered by the sphere may be represented as follows:

$$\vec{E}_1(\vec{r}_0) = \Delta \varepsilon k_0^2 \vec{E}_0 L, \quad (30)$$

where

$$L = \frac{1}{4\pi} \int_V \frac{e^{ik_z(z+R)}}{R} dV. \quad (31)$$

Let us evaluate the integral of L in the spherical system of coordinates. We will consider that the dimensions of the sphere are much smaller than the distance to the point of observation ($a \ll r_0$), that is, we shall examine the field within the wave zone of the sphere. In this case the integral of Eq. (31) may be rewritten as

$$L = \frac{e^{ik_z r_0}}{4\pi r_0} \int_V e^{-ik_z \left[\frac{x_0}{r_0} x + \frac{y_0}{r_0} y - \left(1 - \frac{z_0}{r_0}\right) z \right]} dV, \quad (32)$$

or in spherical coordinates

$$L = \frac{e^{ik_z r_0}}{4\pi r_0} \int_V e^{ik_z 2r \sin \frac{\Theta_0}{2} \cos \xi} dV, \quad (33)$$

where $\cos \xi$ is the cosine of the angle between the lines $\left(\frac{3\pi}{2} + \frac{\Theta_0}{2} \right)$, φ_0 and Θ , φ . Denoting $2 \sin \frac{\Theta_0}{2}$ by b and considering that

$$\frac{1}{4\pi} \int_0^\pi \sin \Theta d\Theta \int_0^{2\pi} e^{ibk_z r \cos \xi} d\varphi = j_0(bk_z r),$$

$$j_0(bk_z r) = \sqrt{\frac{\pi}{2bk_z r}} J_{\frac{1}{2}}(bk_z r);$$

$$J_{\frac{1}{2}}(bk_z r) = \sqrt{\frac{2}{\pi bk_z r}} \sin bk_z r,$$

the integral of Eq. (33) may be finally written in the form

$$L = \frac{e^{ik_z r_0}}{r_0} \int_0^a \frac{\sin bk_z r}{bk_z} r dr, \quad (34)$$

where a is the radius of the sphere.

After integrating in Eq. (34) we obtain

$$L = \frac{e^{ik_2 r_0}}{r_0} \left[\left(\frac{\sin bk_2 a}{bk_2 a} - \cos bk_2 a \right) \frac{a}{b^2 k_2^2} \right]. \quad (35)$$

The magnetic vector of the scatter field is related to the electric vector in the following manner:

$$\vec{H}_1 = \Delta \varepsilon k_0^2 L \sqrt{\varepsilon_2} [\vec{r}^* \vec{E}_0], \quad (36)$$

where \vec{r}^* is the unit vector directed from the coordinate origin to the point of observation.

The power (per unit solid angle) scattered by the sphere in the direction of the observation point is [10]

$$P_\Omega = S_{\vec{r}^*} r_0^2 \quad (37)$$

where $S_{\vec{r}^*}$ is the projection of the vector $\vec{S} = \frac{1}{2} [\vec{E}_1 \vec{H}_1^*]$ in the direction of \vec{r}^* , while the total power scattered by the sphere in all directions,

$$P_s = \int_0^\pi \int_0^{2\pi} P_\Omega \sin \Theta d\Theta d\varphi. \quad (38)$$

Let us define the scatter cross-section Q_s as the ratio of total energy scattered per second to the power flux density of the incident wave, that is,

$$Q_s = \frac{P_s}{S_z}, \quad (39)$$

where $S_z = \frac{1}{2} \sqrt{\varepsilon_2} E_0^2$. Substituting into Eq. (38) the expressions for \vec{E}_1 and \vec{H}_1 from Eqs. (30) and (36) and integrating with respect to Θ and φ , we obtain for the scatter cross-section

$$Q_s = 2\pi a^4 k_2^2 \Delta \varepsilon^2 \left[-\frac{1}{4(2k_2 a)^2} - \frac{1}{8(2k_2 a)^4} + \frac{\sin 4k_2 a}{4(2k_2 a)^3} + \frac{\cos 4k_2 a}{8(2k_2 a)^4} + \frac{1}{4} \right]. \quad (40)$$

Let us examine the two extreme cases:

$$ak_2 \ll 1, \quad (41a)$$

$$ak_2 \gg 1. \quad (41b)$$

Case (41a) corresponds to Rayleigh scattering, case (41b) corresponds to the scattering which occurs in long-distance tropospheric propagation of microwaves (a is of the order of tens of meters, λ is of the order of tens of centimeters).

With $ak_2 \ll 1$ the sine and cosine in square brackets in Eq. (40) may be expanded into a series in powers of $4k_2 a$ and limited to the first four terms of the expansion. After appropriate transposition Eq. (40) is rewritten as

$$Q_s = \pi a^2 \Delta \varepsilon^2 \frac{8}{27} \left(\frac{2\pi a}{\lambda} \right)^4. \quad (42)$$

We have thus derived the well-known Rayleigh formula for the scattering of electromagnetic waves by small particles.

With $ak_2 \gg 1$ the first four terms in square brackets in Eq. (40) may be neglected and only the last term considered. In this case Eq. (40) takes the form

$$Q_s = \pi a^2 \Delta \varepsilon^2 \frac{1}{2} \left(\frac{2\pi a}{\lambda} \right)^2. \quad (43)$$

Using for the scattering cross-section an expression in the form of a series obtained by the usual diffraction method, Jobst [5] obtained closed forms for the scattering cross-section:

$$Q_s = \frac{8}{27} \pi a^2 \Delta \epsilon^2 \left(\frac{2\pi a}{\lambda} \right)^4 \quad (44)$$

for the condition

$$\rho \Delta \epsilon \ll \rho \ll 1 \quad (45)$$

and

$$Q_s = \frac{1}{2} \pi a^2 \Delta \epsilon^2 \left(\frac{2\pi a}{\lambda} \right)^2 \quad (46)$$

for the condition

$$\rho \Delta \epsilon \ll 1 \ll \rho, \quad (47)$$

where $\rho = 2\pi a/\lambda$, wherein inequalities (45) and (47) are equivalent to condition (24). Comparison of Eqs. (42) and (43) with Eqs. (44) and (46) shows their complete identity, which permits the following conclusion: being highly descriptive and mathematically simple, upon fulfillment of condition (24) the perturbation method provides solutions coinciding with the solutions obtained by the exact diffraction method.

B. Let us examine the same problem as in section A, but let us now consider the dielectric constant of the sphere as dependent on r . On the basis of Eq. (28) the expression for the electric vector will have the following form:

$$\vec{E}_1 = k_0^2 \vec{E}_0 \frac{e^{ik_2 r_0}}{r_0} L_1, \quad (48)$$

where

$$L_1 = \int_0^a \epsilon(r) \sin(bk_2 r) r dr. \quad (49)$$

Substituting in Eq. (49) different functions of $\epsilon(r)$, we obtain various dependences of \vec{E}_1 on a , λ and θ . For example, let us examine the relationship, widely used in scatter theory,

$$\epsilon(r) = \Delta \epsilon e^{-\frac{r}{l}}, \quad (50)$$

where

$$l \ll a \quad (51)$$

is a parameter which in scatter theory is considered as a correlation radius. Substituting Eq. (50) in Eq. (49) and integrating, taking inequality (51) into account, we obtain the following expressions for \vec{E}_1 and \vec{H}_1 :

$$\vec{E}_1 = \frac{2\Delta \epsilon k_0^2 l^3}{(1 + b^2 k_2^2 l^2)^2} \frac{e^{ik_2 r_0}}{r_0} \vec{E}_0, \quad (52)$$

$$\vec{H}_1 = \sqrt{\epsilon_2} \frac{2\Delta \epsilon k_0^2 l^3}{(1 + b^2 k_2^2 l^2)^2} \frac{e^{ik_2 r_0}}{r_0} [r^* \vec{E}_0], \quad (53)$$

where $\sqrt{\epsilon_2}$ is the mean characteristic conductivity of the medium. The power per unit solid angle scattered by the sphere in the direction of the observation point is the real part of $\frac{1}{2} [\vec{E}_1 \vec{H}_1^*]$ multiplied by r_0^2 . Then the scatter cross-section Q_s is given by

$$Q_s = \frac{2\pi \Delta \epsilon^2 k_0^4 l^4}{3k_2^2} \left\{ 1 - \frac{1}{(1 + 4k_2^2 l^2)^3} \right\}. \quad (54)$$

For the special cases:

$$(1) \quad k_2 l \ll 1,$$

$$Q_s = 8\pi l^2 \Delta \epsilon^2 \left(\frac{2\pi l}{\lambda} \right)^4, \quad (55)$$

$$(2) \quad k_2 l \gg 1,$$

$$Q_s = \frac{2\pi}{3\epsilon_2} l^2 \Delta \epsilon^2 \left(\frac{2\pi l}{\lambda} \right)^2. \quad (56)$$

From Eqs. (55) and (56) it follows that scattering at a dielectric sphere of radius a with variable ϵ under condition (50) may be regarded as scattering at a sphere of radius l with constant ϵ .

3. RELATION BETWEEN THE CORRELATION FUNCTION $R(r)$ USED IN SCATTER THEORY AND THE FUNCTION $\epsilon(r)$

It is of interest to determine that dependence of the dielectric constant of a sphere $\epsilon(r)$ such that at the receiving point there is created a field equal to the field which would result from the scattering of radio waves at random discontinuities of the troposphere.

The complex power density at the receiver, resulting from scatter within the volume V , will be [11]

$$\frac{1}{2} [\vec{E}_1 \vec{H}_1]_{r^*} = \frac{1}{2} V \epsilon_2 E_0^2 \left\{ \frac{k_0 \sin \chi}{4\pi r_0} \right\}^2 4\pi \frac{V \Delta \epsilon^2}{b k_2} \int_0^r R(r) \sin(b k_2 r) r dr, \quad (57)$$

where $R(r)$ is the correlation function of the dielectric constant of air; V is the scatter volume. The power density resulting from scattering of a plane electromagnetic wave at a dielectric sphere with variable $\epsilon(r)$ is

$$\frac{1}{2} [\vec{E}_1 \vec{H}_1]_{r^*} = \frac{1}{2} V \epsilon_2 E_0^2 \left\{ \frac{k_0^2 \sin \chi}{r_0} \right\}^2 \frac{\Delta \epsilon^2}{b^2 k_2^2} \left[\int_0^r \epsilon(r) \sin(b k_2 r) r dr \right]^2. \quad (58)$$

Comparing Eqs. (57) and (58), we obtain the following equation for determining the relation between $R(r)$ and $\epsilon(r)$:

$$\frac{b k_2 V}{4\pi} \int_0^r R(r) \sin(b k_2 r) r dr = \left[\int_0^r \epsilon(r) \sin(b k_2 r) r dr \right]^2. \quad (59)$$

Extracting the square roots and differentiating Eq. (59) with respect to r , we obtain the following relation between $\epsilon(r)$ and the correlation function $R(r)$:

$$\epsilon(r) = R(r) \sqrt{\frac{b k_2 V}{16\pi \int_0^r R(r) \sin(b k_2 r) r dr}}. \quad (60)$$

From Eq. (60) we may draw the following interesting conclusion.

The scattering of radio waves at statistical discontinuities of the troposphere may be regarded as scattering at a sphere at the center of which the dielectric constant is $\sqrt{\frac{V}{\Delta \epsilon^2}}$

and decreases toward its extremities according to the law expressed in Eq. (60), while its dimensions correspond to the dimension l of the correlation variation. In the special case where $\epsilon(r)$ is given by Eq. (50) the scattering of radio waves at statistical discontinuities of the troposphere may reduce to scattering at a sphere of radius l with fixed dielectric constant [11]. In other words, the problem of the scattering of radio waves at statistical discontinuities of the troposphere may be reduced to the problem of the scattering of radio waves at a sphere whose dielectric constant is a regular function associated with the

correlation function by Eq. (60).

In conclusion the author expresses his thanks to B.A. Vvedenskiy, A.V. Sokolov and N.A. Armand for review and discussion of the present work and their valuable comments and advice.

REFERENCES

1. H. Staras, J. Appl. Phys., 1952, 23, 10, 1152.
2. A.D. Wheelon, Proc. IRE, 1955, 43, 10, 1459.
3. G. Eckart, Compt. rend., 1957, 244, 25, 3044.
4. D.G. Paul, IRE Trans., 1958 AP-6, 1, 61.
5. G. Jobst, Ann. Physik, 1925, 78, 18, 157.
6. V.I. Smirnov, Course in Higher Mathematics, IV, GITTL, 1957.
7. G.N. Watson, Theory of Bessel Functions, Part I, IL, 1949.
8. E.V. Hobson, Theory of Spherical and Elliptical Functions, IL, 1952.
9. L.A. Chernov, Propagation of Waves in a Medium with Random Discontinuities, Izd. AN SSSR, 1958.
10. D.A. Stratton, Electromagnetic Theory, GTI, 1948.
11. H.G. Booker, W.E. Cordon, Proc. IRE, 1950, 38, 4, 401.

Submitted to the editors 14 December 1950

GRAPHICAL METHOD OF DESIGNING APLANATIC ANTENNAS

N.G. Ponomarev

The paper describes a simple method of designing reflector antennas and lens antennas satisfying the Abbe sine condition. This method is based on geometrical optics and is similar to the well-known method of designing mirrors from a diagram of special shape. The results of an experimental investigation of a two-reflector aplanatic antenna are given.

In antenna engineering both reflector and lens antennas are used, wherein beam scanning is achieved by displacement of the primary source of radiation over a certain curve. As is known, undistorted beam scanning is possible upon satisfying the "sine condition," which in optics provides undistorted images of objects of finite extent. By analogy with optical systems, antennas in which the sine condition is fulfilled are known as aplanatic antennas.

Reflector and lens systems in which the sine condition is fulfilled have two reflecting (refracting surfaces, since, in addition to the sine condition, they must satisfy the "cophasality condition" ensuring the transformation of a beam of rays emanating from a point into a parallel beam.

Equations for the curves of the central section of an aplanatic reflector system are known [1], but they are complex and calculation with them is extremely laborious. Profiles of a double-refracting aplanatic lens may be plotted as the result of integration of the differential equation [2]. However, the analytic solution of this equation is unknown.

Below we present a simple graphical method of designing both reflector and lens aplanatic systems. It is based on methods of geometrical optics and is similar to the method of plotting the curve of the central section of a mirror from given diagrams of a primary source and a mirror [3].

In order to fulfill the cophasality condition it is necessary that all the rays leaving the antenna be parallel to its axis. The sine condition is fulfilled if all the incident rays from the focus and their corresponding emergent rays intersect at a circle whose center lies at the focus of the system [2]. The radius of this circle is known as the focal length.

In order to plot the profile of a reflector system satisfying both conditions let us draw a circle of radius f with its center at the coordinate origin (see Fig. 1). The coordinate origin is taken as the focus and line Fx as the axis of the system. We plot a system of incident rays from the focus and their corresponding emergent rays parallel to the axis and intersecting the incident rays at the circle. At the outermost ray and the corresponding emergent ray let us select arbitrary points A and B and join them by a straight line. These points are the extreme points of the profile of the reflector system*. Selecting points A and B , as shown in Fig. 1, and joining them by a straight line, we thereby define the path of the extreme ray. In order for the ray to have the shape of the broken line $FABD$ the tangents to the surfaces of the reflectors at points A and B must be perpendicular to the bisectors of angles α and β (see Fig. 1). Plotting these tangents, we find the position of the elementary plane reflectors at the given points. Extending the tangents to the reflector profiles at points A and B to intersection at points A_1 and B_1 with the following incident rays from the focus and their corresponding emergent rays, and joining points A_1 and B_1 by a straight line, we obtain the path of the adjacent ray. Drawing at points A_1 and B_1 straight-line segments, the normals to which coincide with the bisectors of angles α_1 and β_1 , we obtain the next elements of the reflector profiles. Subsequent elements of the profile are obtained by continuing the plotting in the same manner.

As the result of plotting we obtain the broken lines $AA_1A_2A_3 \dots A_n$ and $BB_1B_2B_3 \dots B_n$ approximating the profiles of the reflectors of a system satisfying the cophasality and sine conditions. It is evident that the smaller the angle $\Delta\varphi$, the more accurate the approximation. The accuracy of the graphical plotting may be increased by other means, for example, by the method used in the calculation of lenses with a given amplitude-phase distribution at the aperture [5]. Since the system is symmetrical relative to the focal axis, plotting may be limited to the profile of only one side of the axis. The second half is obtained by reflection relative to the axis.

Figure 2 shows the plot of the profiles of a two-reflector system for the case where point A lies outside the circle of radius f and point B lies within it.

On the basis of graphical plotting we may obtain recursion formulas expressing the coordinates of points A_{k-1} and B_{k-1} . For the two-reflector system these formulas have the form

$$x_{A_k} = x_{A_{k-1}} \frac{\operatorname{tg} [\varphi_{\max} - (k-1) \Delta\varphi] + \operatorname{ctg} \left[\frac{\varphi_{\max} - (k-1) \Delta\varphi + \beta_{k-1}}{2} \right]}{g [\varphi_{\max} - k \Delta\varphi] + \operatorname{ctg} \left[\frac{\varphi_{\max} - (k-1) \Delta\varphi + \beta_{k-1}}{2} \right]}, \quad (1)$$

*The selection of initial points A and B is somewhat arbitrary. It is not difficult to see that both points may lie within a circle of radius f or outside it or that one of them (A or B) may lie within it and the other outside it. One or another choice of initial points determines the shape of the profile of the reflectors and their relative position.

$$y_{A_k} = x_{A_k} \operatorname{tg}(\varphi_{\max} - k\Delta\varphi), \quad (2)$$

$$x_{B_k} = x_{B_{k-1}} + 2f \operatorname{tg} \frac{\beta_{k-1}}{2} \sin \frac{\Delta\varphi}{2} \cos \left[\varphi_{\max} - \frac{(2k-1)\Delta\varphi}{2} \right], \quad (3)$$

$$y_{B_k} = f \sin(\varphi_{\max} - k\Delta\varphi), \quad (4)$$

where

$$\operatorname{tg} \beta_{k-1} = \frac{y_{B_{k-1}} - y_{A_{k-1}}}{x_{B_{k-1}} - x_{A_{k-1}}}, \quad (5)$$

$k = 1, 2, 3, \dots, n$.

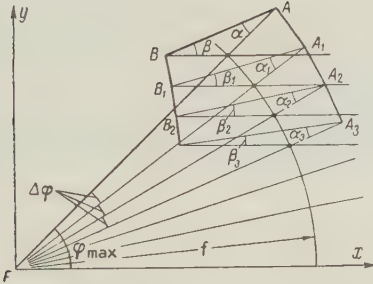


Fig. 2. Plot of the profile of a two-reflector aplanatic system (point A is outside the circle of radius f and point B is within it).

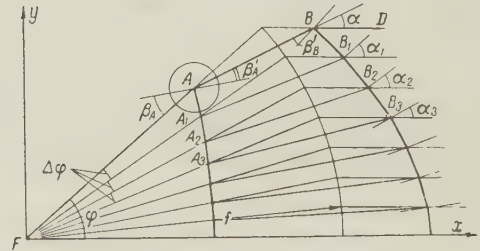


Fig. 3. Plot of the profile of an aplanatic lens.

In deriving Eqs. (1) - (5) it was assumed that φ_{\max} is divided into identical angles $\Delta\varphi$.

Let us discuss the graphical plot of the refractive profiles of an aplanatic double-refracting lens. For this purpose, as in plotting a two-reflector system, let us plot a family of the rays incident from the focus and the corresponding rays emergent from the lens — intersecting at a circle of radius f (see Fig. 3). Joining points A and B by a straight line, let us assume that the outermost ray has the shape of broken line FABD. In addition, if the refractive index of the lens is $n > 1$, point A is chosen on a ray incident from the focus within the circle, while B is chosen on an emergent ray outside the circle*. Since the refractive index of the lens is considered known, by means of the law of refraction we can determine the position of the elementary refractive surfaces (at points A and B) ensuring the given shape of a ray in the form of the broken line FABD. It is not difficult to determine that the angles of incidence β_A and β_B at points A and B are, respectively,

$$\beta_A = \operatorname{arctg} \frac{-n \sin \alpha_A}{1 + n \cos \alpha_A}, \quad (6)$$

$$\beta_B = \operatorname{arctg} \frac{n \sin \alpha}{n \cos \alpha - 1}, \quad (7)$$

where α_A is the large angle between lines FA and AB and α is the small angle between lines AB and BD.

The angles of incidence β_A and β_B , and consequently also the normal to the elementary refractive surfaces at points A and B, can be plotted graphically. Let us discuss the plot of

*It can be shown by graphical plotting that for an aplanatic dielectric lens ($n > 1$) point A must lie within the circle of radius f , while B must lie outside it. For a lens with $n < 1$ points A and B must lie within this circle.

the normal to the refractive surface at point A. Let us plot a unit circle (see Fig. 4) with its center at point A and on line AB let us mark off segment AN', the length of which is equal to n. From the point N' we drop a perpendicular to the line FA and draw a line from N' to M', which is obtained as a result of the intersection of the unit circle with the extension of the line FA. As is seen from the figure and from Eq. (6), the angle N'M'P is equal to the angle of incidence β_A at point A, while line M'N' is parallel to the normal to the profile of the refractive surface. Thus, the position of the elementary refractive surface at point A has been determined. Plotting of the normal at point B and at all subsequent points A_1, A_2, A_3, \dots and B_1, B_2, B_3, \dots is performed in the same manner as at point A. The entire procedure for plotting the profiles of the refractive surfaces of a lens is the

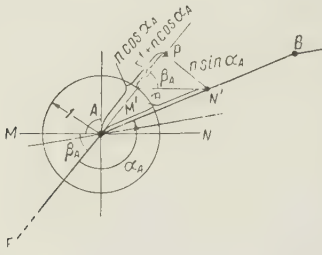


Fig. 4. Graphical plot of the normal MN and the angle of incidence β_A at point A.

same as in the case of a two-reflector system (see Fig. 3).

As the result of the plotting we obtain the approximate shape of the profiles of the refractive surfaces of an aplanatic lens. The recursion formulas for the coordinates of the points of the lens profile have the form

$$x_{A_k} = x_{A_{k-1}} \frac{\operatorname{tg} [\varphi_{\max} - (k-1) \Delta\varphi] + \operatorname{ctg} (\alpha_{k-1} - \beta'_{A_{k-1}})}{\operatorname{tg} (\varphi_{\max} - k \Delta\varphi) + \operatorname{ctg} (\alpha_{k-1} - \beta'_{A_{k-1}})}, \quad (8)$$

$$y_{A_k} = x_{A_k} \operatorname{tg} (\varphi_{\max} - k \Delta\varphi), \quad (9)$$

$$x_{B_k} = x_{B_{k-1}} + 2f \operatorname{tg} \beta_{B_{k-1}} \sin \frac{\Delta\varphi}{2} \cos \left[\varphi_{\max} - \frac{(2k-1) \Delta\varphi}{2} \right], \quad (10)$$

$$y_{B_k} = f \sin (\varphi_{\max} - k \Delta\varphi), \quad (11)$$

where

$$\operatorname{tg} \beta_{B_{k-1}} = \frac{n \sin \alpha_{k-1}}{n \cos \alpha_{k-1} - 1}; \quad (12)$$

$$\operatorname{tg} \beta'_{A_{k-1}} = \frac{\sin [\varphi_{\max} - (k-1) \Delta\varphi - \alpha_{k-1}]}{n - \cos [\varphi_{\max} - (k-1) \Delta\varphi - \alpha_{k-1}]}; \quad (13)$$

$$\operatorname{tg} \alpha_{k-1} = \frac{y_{B_{k-1}} - y_{A_{k-1}}}{x_{B_{k-1}} - x_{A_{k-1}}}. \quad (11)$$

Plotting the profiles both of the two-reflector and lens antennas was performed by beginning with the outermost points. However, such plotting may be performed, by assuming that the initial points lie on the axis, and being given the thickness of the lens or the distance between reflectors. Such plotting is preferred for lenses, for a lens of minimum thickness may be derived in this manner.

Several types of aplanatic systems were developed by means of this method. A few profiles of reflector and lens antennas are given in Fig. 5-8. As is seen from these figures, it may develop that one of the reflectors considerably shields the other (see Fig. 6). However, in actual construction of a two-reflector system such shielding may be avoided by various means. For example, if for a given antenna the sine condition is satisfied in the horizontal plane, it is necessary to displace the reflectors in the vertical plane (see Fig. 9). Then there will exist no shielding between reflectors.

The above method was used in designing a reflector antenna the experimental model of which is shown in Fig. 9. Scanning of the beam is achieved by the rotation of a parabolic horn, which for this purpose is mounted on a special bracket. The ratio of the focal length f to the aperture D was 0.8. Figure 10 shows experimental radiation patterns which were plotted by scanning of the pattern maximum by $\pm 50^\circ$ from the antenna axis (in Fig. 10 the

angular values are referred to the width of the pattern in the central position). As seen from Fig. 10 the experimental model of the antenna (with $f/D = 0.8$) ensures undistorted scanning over a wide angular sector. The accuracy of the plotting of the reflector profiles was adequate for practical purposes.

The author is indebted to L.S. Benenson for his interest in this work and his valuable advice.

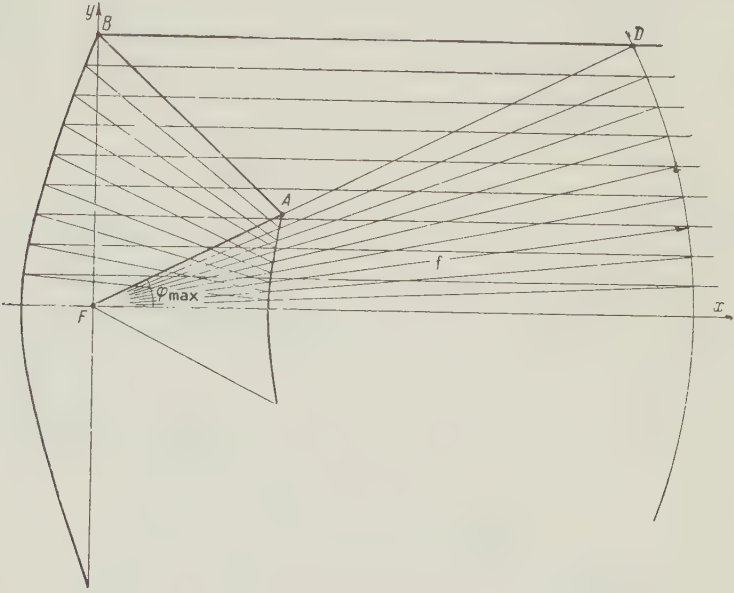


Fig. 5. Profiles of the reflectors of an aplanatic system with negligible shielding of the large reflector by the small.

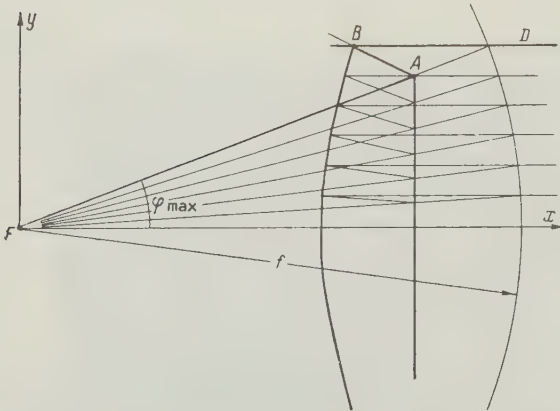


Fig. 6. Profiles of the reflectors of an aplanatic system in which, due to considerable shielding of one reflector by the other, it is convenient to displace the reflectors vertically.

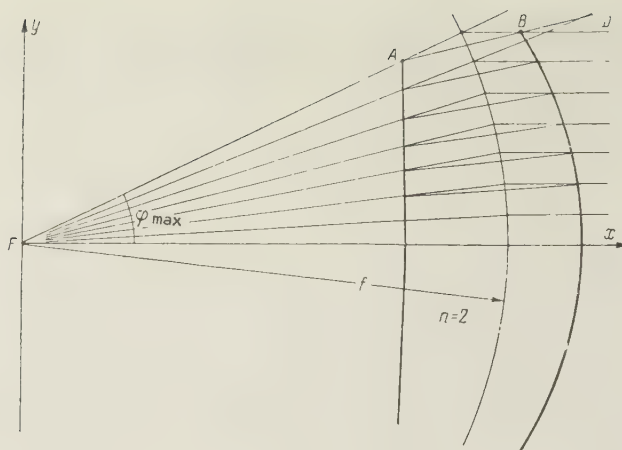


Fig. 7. Profiles of an aplanatic lens.

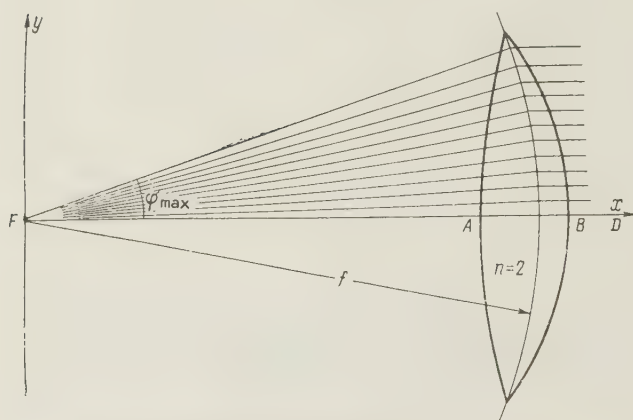


Fig. 8. Profiles of a "thin" aplanatic lens.



Fig. 9. Experimental model of an antenna.

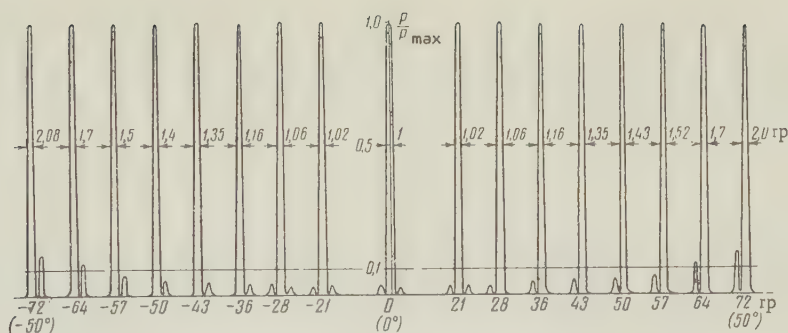


Fig. 10. Radiation pattern of two-reflector system.

REFERENCES

1. K. Schwartzchild, Investigations in Geometrical Optics, Parts I, II, 1905, translation 257, BNT MO, 1946.
2. James J. DeVor, S. Brown, Jones, et al., Lens antennas with scanning beams, Sb. Vopr. radiolokatsionnoy tekhniki, 1951, No. 5, pp. 3-37.
3. A.S. Dunbar, Calculation of doubly curved reflectors for shaped beams, Proc. IRE, 1948, 36, 10, 1289.
4. Ya. N. Fel'd, N.S. Benenson, Antenna Feeder Installations, Part II, 1959, Izd. VVIA im. N. Ye. Zhukovskiy.
5. G.P. Foldes, L. Solymar, Lens aerial design, Electronic and Radio Engineering, 1959, 36, II, 2, 73.

Submitted to the editors 28 May 1960

SUPPRESSION OF CURRENTS EXCITED IN METAL SHIELDS BY DIFFRACTION ANTENNAS OF FINITE DIMENSIONS

O.N. Tereshin, A.Ye. Sokolov

The article discusses a method of suppressing electrical currents excited by a diffraction antenna at a metal surface. Suppression is achieved by means of a purely reactive surface impedance. The method of suppression may be used for the decoupling of two diffraction antennas as well as for lowering the back radiation level and decreasing the effect of metal

shield dimensions on the radiation pattern of diffraction antennas. Experimental data show that the degree of suppression of surface currents is determined by the maximum rate of change of the purely reactive surface impedance which is attained in practical realization of the latter.

INTRODUCTION

Reference [1] discussed the possibility of creating a decoupling arrangement between two slot antennas of unlimited length. It was shown that in this two dimensional case the degree of decoupling is determined by the maximum attainable rate of change of the purely reactive surface impedance. Reference 2 pointed out that in the creation of decouplers for antennas of finite length the attainable degree of decoupling is limited by the level of the currents surrounding a decoupling structure built on the basis of the conclusions obtained from an examination of the two dimensional problem.

This paper presents the result of theoretical and experimental investigations of decouplers for annular-slot antennas of finite dimensions. Results are given for an experimental check of the operation of decouplers constructed in accordance with the conclusions of Ref. 1 as well as of the effect of decouplers on the radiation patterns of slot antennas.

1. STATEMENT OF THE PROBLEM

Let us examine the problem of a decoupling arrangement for diffraction antennas of finite dimensions.

An impedance surface is given in the plane $z = 0$.

In the region S at the coordinate origin there is located a diffraction antenna (Fig. 1) in which the current distribution is symmetrical with respect to φ . In this case the boundary conditions at the impedance surface with $z = 0$ may be written in the form

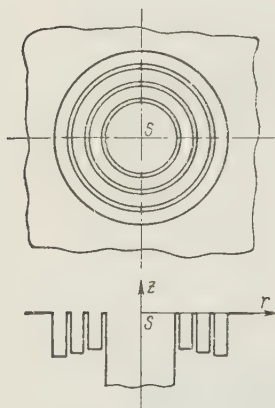


Fig. 1

$$\begin{aligned} Z(r) &= -\frac{E_r}{H_\varphi} \text{ with } z = 0, \\ \frac{E_\varphi}{H_r} &= 0 \quad \text{with } z = 0, \end{aligned} \quad (1)$$

where E_r , E_φ , H_r , H_φ are the tangential components of the electric and magnetic vectors of the field; $Z(r)$ is the distribution function of the surface impedance.

In obtaining a purely reactive surface impedance by means of a ribbed structure Eq. (1) is applicable in the case where the width of the channels is much smaller than the wavelength and the width of the ribs is much smaller than the channel width.

Let us assume that the field of the primary sources given in region S has the free-space properties of TM waves.

As shown in Ref. 1, the degree of limitation of current spread may be characterized by the degree of diminution of the tangential component of the magnetic field at the surface $z = 0$ in the radial direction. The solution is similar to the calculations performed in Ref. 1; hence we shall discuss only the principal stages. As before, we are given the field in the form

$$H_\varphi(r, z) = e^{-Z^*(r, z)}, \quad (2)$$

where

$$Z^*(r, z) = Z_1(r, z) + iZ_2(r, z). \quad (3)$$

Here Z_1 and Z_2 are real functions of the arguments r and z .

For the special case of TM waves from the primary source we obtain from Maxwell's

equations

$$E_r = -\frac{i}{\omega\epsilon} \frac{\partial H_\phi}{\partial z}, \quad (4)$$

where ϵ is the absolute dielectric constant of the medium.

Substituting Eqs. (4) and (2) into Eq. (1), we obtain

$$Z(r) = -\frac{i}{\omega\epsilon} \frac{\partial Z''}{\partial z} = -\frac{i}{\omega\epsilon} \frac{\partial Z_1}{\partial z} + \frac{1}{\omega\epsilon} \frac{\partial Z_2}{\partial z} \text{ with } z = 0. \quad (5)$$

The condition for a purely reactive surface impedance may be written in the form

$$\frac{\partial Z_2}{\partial z} = 0 \text{ with } z = 0. \quad (6)$$

The tangential component H_ϕ of the magnetic field must satisfy the homogeneous wave equation

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial H_\phi}{\partial r} \right) - \frac{H_\phi}{r^2} + \frac{\partial^2 H_\phi}{\partial z^2} + k^2 H_\phi = 0. \quad (7)$$

Further, as in Ref. 1, in order to determine the class of functions satisfying the boundary conditions (1) and the wave equation (7) it suffices to consider the region adjacent to the plane $z = 0$ where the functions Z_1 and Z_2 may be written as asymptotic expansions in powers of z in the form

$$\begin{aligned} Z_1(r, z) &= Z_1^0(r) + Z_1'(r)z + Z_1''(r)z^2, \\ Z_2(r, z) &= Z_2^0(r) + Z_2'(r)z + Z_2''(r)z^2. \end{aligned} \quad (8)$$

Substituting the asymptotic expansions (8) into the wave equation (7) and assuming z to be sufficiently small and then separately equating the real and imaginary parts to zero, we arrive at the equations

$$Z_1'(r) = \pm \sqrt{\left(\frac{dZ_2^0}{dr}\right)^2 - \left(\frac{dZ_1^0}{dr}\right)^2 - \frac{1}{r} \frac{dZ_1^0}{dr} - \frac{d^2 Z_1^0}{dr^2} - 2Z_1''(r) - \left(k^2 - \frac{1}{r^2}\right)}, \quad (9)$$

$$\frac{d^2 Z_2^0}{dr^2} - \frac{1}{r} \frac{dZ_2^0}{dr} + 2Z_2''(r) + 2\left(\frac{dZ_1^0}{dr}\right)\left(\frac{dZ_2^0}{dr}\right) = 0. \quad (10)$$

For the case of diffraction antennas it is necessary to assume in Eq. (8) that Z_1'' and Z_2'' are equal to zero.

Using the equations derived in Ref. 3, it is easily shown that at those points of the surface $z = 0$ lying within the region S where the primary diffraction antenna is located it is necessary to choose the minus sign in Eq. (9) in order to fulfill the radiation conditions.

With the limitations presented, the condition for a purely reactive surface impedance is written in the form

$$\frac{d^2 Z_2^0}{dr^2} - \frac{1}{r} \frac{dZ_2^0}{dr} + 2\left(\frac{dZ_1^0}{dr}\right)\left(\frac{dZ_2^0}{dr}\right) = 0, \quad (11)$$

while the required distribution of surface impedance is found from the equation

$$Z(r) = \frac{i}{\omega\epsilon} \sqrt{\left(\frac{dZ_2^0}{dr}\right)^2 - \left(\frac{dZ_1^0}{dr}\right)^2 - \frac{1}{r} \frac{dZ_1^0}{dr} - \frac{d^2 Z_1^0}{dr^2} - \left(k^2 - \frac{1}{r^2}\right)}. \quad (12)$$

If the coordinate of the observation point (r) in Eqs. (11) and (12) tends to infinity, then these expressions asymptotically become the expressions derived in Ref. 1 for the two

dimensional problem.

Thus, the distributions of purely reactive surface impedance which have been derived for one and the same variations in the tangential component of the magnetic field about the surface $z = 0$ differ in the case of a symmetrical antenna of finite dimensions and the two dimensional case only at small distances from the antenna.

3. EXPERIMENTAL RESULTS

(a) Measurement of the distribution of the amplitude and phase of the tangential component of the magnetic field along the structure. Experimental investigation of the distribution of the amplitude and phase of the surface currents excited by a slot antenna was performed by the usual method on the prototype shown in Fig. 2. The required purely reactive surface impedance was attained by means of a ribbed structure with period $b = 0.156\lambda$ and rib thickness $\Delta = 0.03\lambda$. The length of the attenuating structure was 1.56λ . In the calculation of the structure, the function for the change in amplitude of the tangential component of magnetic field H_x along the y -axis was given in the form

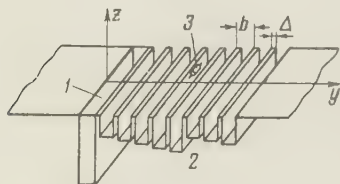


Fig. 2. 1) Slot antenna; 2) supporting structure; 3) measuring probe.

$$|H_x| = e^{Z_1^0(y)} = e^{-\beta y}, \quad (13)$$

where β is the attenuation constant. Then the required variation in the purely reactive surface impedance, in accordance with Ref. 1, is defined by

$$Z_s(y) = \frac{i}{\omega \epsilon} \sqrt{e^{4\beta y} C - \beta^2 - k^2}. \quad (14)$$

Results of the experiment are shown in Fig. 3 and Table 1. Figure 3 gives the calculated (a) and experimental (b) distributions of amplitude and phase (for $\beta = 0.475 k$) of the field along the attenuating structure for various attenuation constants. The rectangles represent the channel depths of the ribbed structure for the case where $\beta = 0.475 k$.

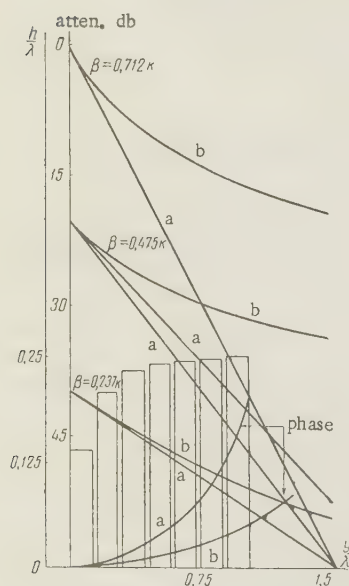


Fig. 3

Table 1 lists the values of field attenuation caused by the entire suppressing structure.

An investigation of the suppression of surface currents excited by a primary antenna in the form of an open-ended circular waveguide (E_{01} mode) within a metal shield was performed on the model shown in Fig. 4.

The required purely reactive surface impedance was attained by means of a ribbed-disc structure with period $b = 0.12\lambda$ and $\Delta = 0.01\lambda$.

The variation of the amplitude of the tangential component H_ϕ of the magnetic field along the radial axis was given in the form

$$|H_\phi| = \frac{e^{-\beta r}}{r}. \quad (15)$$

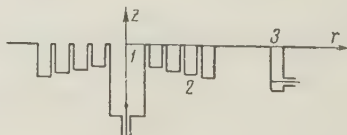


Fig. 4. 1) Transmitting antenna; 2) suppressing structure; 3) receiving (slot) antenna.

Table 1

| Attenuation constant, β | Coefficient C | Design attenuation, db | Measured attenuation, db |
|-------------------------------|---------------|------------------------|--------------------------|
| 0.237k | $1.05k^2$ | 20 | 15 |
| 0.475k | $1.22k^2$ | 40 | 14 |
| 0.712k | $1.53k^2$ | 60 | 19 |

Table 2

| Attenuation constant, β | Design attenuation, db | Measured attenuation, db |
|-------------------------------|------------------------|--------------------------|
| 0.16k | 8.4 | 8.0 |
| 0.32k | 16.7 | 14.0 |
| 0.48k | 25.1 | 16.0 |
| 0.64k | 33.5 | 16.0 |

The required variation of the surface impedance, according to Eq. (12), is determined from

$$Z(r) = \frac{i}{\omega\epsilon} \sqrt{C \frac{e^{4\beta r}}{r^3} - \beta^2 - \frac{\beta}{r} - \left(k^2 - \frac{1}{r^2}\right)}. \quad (16)$$

The measurement of the amplitude of the surface currents was performed on the basis of the degree of excitation of a receiving antenna (slot) placed a fixed distance away from the transmitting antenna. A suppressing structure of length 0.9λ was introduced into the zone between the receiving and transmitting antennas.

Results of the tests are given in Table 2 (only the additional attenuation attributable to the presence of the suppressing structure is listed).

Let us evaluate the results.

Examination of the results shows that in both cases there is adequate correspondence between the design and experimental values of field attenuation and the variation in amplitude of the field along the structure for small values of the attenuation constant β . The deviation in results for large values of the attenuation constant may be explained as follows. With an increase in the attenuation constant there is also required a higher rate of change in the purely reactive surface impedance [see Eqs. (14), (16)], which condition cannot be realized by means of a ribbed structure with finite channel width. Actually, due to the finite channel width and wall thickness the required continuous distribution of purely reactive surface impedance is approximated in steps. Hence, in practice the variation in the purely reactive impedance is determined by the first 3-4 channels, after which the impedance ceases to increase due to inaccuracies in channel depth (see Fig. 3). With this in mind, it is convenient, in order to increase the attenuation, to periodically repeat over the entire structure the variation in surface impedance as produced by the first channels.

If we extend over the entire structure the variation reflected in the first three channels for the case of a slot antenna and $R = 0.475k$, then the attenuation increases considerably and reaches 33 db (Curve d in Fig. 3). For the disc structure (repeating over the entire structure the variation reflected in the first two channels) the following results are obtained: with $\beta = 0.48k$ the attenuation rose to 25 db, with $\beta = 0.64k$ the attenuation rose to 33 db.

With given dimensions of the suppressing structure the degree of attenuation may be increased by using a ribbed structure with small period ($b \approx 0.025\lambda$ and less). In testing for the case of a disc with purely reactive surface impedance, such a structure was made in the form of dielectric rings of different height wrapped in brass foil. The structure was also made of such rings in close contact with one another.

Measurements showed that with such a structure of size 0.1λ , the resulting decoupling was 10 db, which corresponds to $\beta = 2.4k$. It is obvious that this is not the maximum value attainable. The use of ribbed structures with still smaller periods will result in considerably greater attenuation of the field.

We note also that in periodic repetition of the segment with greatest rate of change in purely reactive surface impedance Eqs. (14) and (16) become inapplicable to the segment at a sudden reverse jump in surface impedance (transition to the next period) as soon as at this point $dZ_1/dr \rightarrow \infty$, and as shown in Ref. 1, Eqs. (14) and (16) are applicable with accuracy up to terms of the order $\frac{dZ_1}{dr} z$.

However, the curve for the change in amplitude of the field with periodic repetition of segments with sudden change in the purely reactive surface impedance (Fig. 3, Curve d) shows that the variation in amplitude of the field is quite close to that calculated for a

continuous change. This means that within the limits of each period of change in impedance Eqs. (14) and (16) remain applicable and at each point of jump in the surface impedance the coefficient of reflection is small.

Thus, in the final analysis the rate of change in the purely reactive surface impedance can be determined.

(b) Use of a suppressing structure in eliminating the influence of a metallic surface (beyond the limiting structure) on the radiation pattern of a slot antenna. For this purpose the radiation patterns of two models which differ in the presence or absence of a metal shield beyond the suppressing structure were plotted. The suppressing structure had a length of 0.1λ and provided attenuation on the order of 10 db.

Figure 5 shows the radiation patterns of a narrow slot antenna cut out of the center of a shield with length $6\lambda_0$ (the schematic representation of each of the experimental models

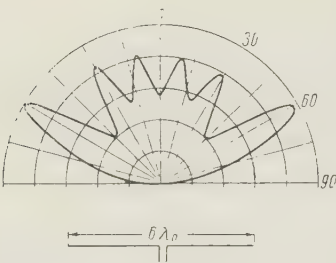


Fig. 5

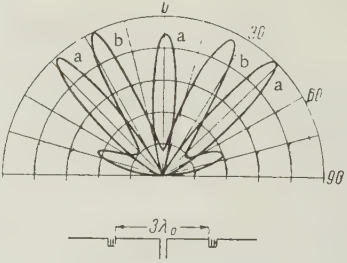


Fig. 6

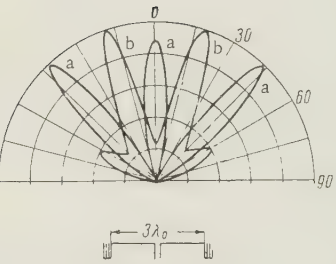


Fig. 7

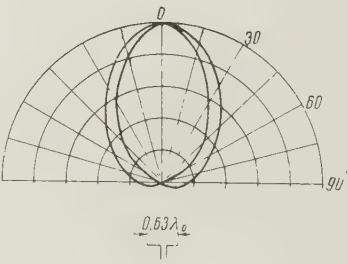


Fig. 8

is shown beneath its corresponding radiation pattern in each figure). Figure 6 shows the radiation patterns of the same slot antenna with an attenuating structure at the center of each arm, designed for two frequencies differing by 20% (a for frequency f_0 and b for frequency $f_1 = 0.8f_0$). Figure 7 shows the radiation pattern for the same frequencies in the case where the outer shield portions are omitted. The patterns in Figure 8 are plotted from a model with a short shield ($L = 0.63\lambda_0$) for the case of the omission of the attenuating structure at the end of the shield (wider pattern) and for the case of its use (narrower pattern). In comparing the radiation patterns for frequency f_0 in Figures 6 and 7, it is evident that elimination of the outer portions of the shield does not distort the radiation pattern. This indicates the absence of currents in these portions.

The radiation pattern in this case is created both by the radiation from currents flowing along the shield and by the radiation from the suppressing structure (compare with Fig. 5), which is confirmed by the change in radiation pattern at a frequency f_1 differing from f_0 by 20%. At frequency f_1 a slight distortion is observed upon eliminating the shield sections beyond the suppressing structure. This is explained by a decrease in the cut-off properties of the ribbed structure because of the decrease in the drops in purely reactive surface impedance along the attenuating structure with a decrease in frequency.

Installation of the suppressing structure at the edge of a small shield (Fig. 8) leads to a decrease in the back-radiation level.

CONCLUSIONS

1. In practice, an adequate realization of the required purely reactive surface impedance results in an excellent agreement of the surface current attenuation with calculated values.

2. For the purpose of providing maximum attenuation, optimum results are obtained by the use of periodically repeated segments with sharply variable purely reactive surface impedance.

3. A suppressing structure in the form of periodically repeated segments of ribbed structures with a small period permits considerably greater decoupling than is obtainable with known decoupling devices of the "cut-off slot" type [4] with identical dimensions (27 db in comparison with 9 db). An even greater degree of decoupling is provided by decreasing the period of the ribbed structure.

4. The use of a suppressing structure in the form of periodically repeated segments with sharply variable purely reactive surface impedance along its length permits eliminating the influence of a metallic surface beyond the limiting structure of the radiation pattern of a slot (diffraction) antenna.

5. The use of limiting structures at the edge of the shield of a slot antenna reduces the back-radiation level.

6. A ribbed structure with purely reactive surface impedance loses little of its limiting properties with a change of 20% in frequency. The changes in radiation patterns at different frequencies are apparently associated with the change in nature of the radiation of the limiting structure as well as with the change in current distribution in the segment from the antenna to the limiting structure.

It is to be expected that by an appropriate choice of the radiation of the limiting structure it will be possible to establish optimum variations in the purely reactive surface impedance for a limiting structure with a view to small distortions of the radiation pattern due to a shield of finite dimensions.

REFERENCES

1. O.N. Tereshin, Decoupling of two slot antennas by means of a ribbed structure in the slot plane, *Radiotekhnika i elektronika*, 1960, 5, 12, 1944.
2. O.N. Tereshin, A.S. Belov, Decoupling of slot antennas by means of an impedance structure in the slot plane, *Izv. Vuzov MVO SSSR (Radiotekhnika)*, 1960, 3, 3, 359.
3. G.T. Markov, Excitation of a circular waveguide, *ZhTF*, 1952, 22, 5, 747.
4. W.K. Saunders, Control of surface currents by the use of channels, *IRE Trans.*, 1956, AP-4, 1.

Submitted to the editors 11 June 1960

TRANSITION CALCULATIONS FOR A SYMMETRICAL MAGNETIC WAVE IN A CIRCULAR WAVEGUIDE

B. Z. Katsenelenbaum, Z. A. Malina

The paper discusses methods of numerical calculation of a symmetrical waveguide transition between two waveguides of different diameter for a symmetrical magnetic wave.

INTRODUCTION

In the transition of an H_{01} wave in a circular waveguide from a waveguide of one diameter to a waveguide of another diameter there arises reflection of the H_{01} wave and parasitic waves H_{0i} ($i > 1$) in both directions. In a number of cases the occurrence of parasitic waves substantially disturbs the operating conditions of the waveguide system. A decrease in the amplitude of the parasitic waves over a wide range of frequencies by increasing the length of the tapered transition requires the use of extremely long (more than 1 m) tapers. Below we present several methods for the calculation of tapered transitions with generatrices of complex shape. The use of such tapered transitions permits obtaining small transformation with a relatively small length of transition. The calculation is based on the system of differential equations for wave amplitudes within a nonuniform waveguide as obtained in Ref. 1 and discussed in Refs. 2 and 3.

The presence of the so-called critical section for a parasitic wave introduces a complication in the mathematical organization of the problem. The method developed in Ref. 4 is used in the analysis of this case. The examples given in this paper illustrate the particular characteristics of the proposed methods of calculation.

1. SELECTION OF GENERATRIX

In choosing the shape of a tapered section we shall proceed from the approximate analogy (see Refs. 2 and 5) existing between the problem discussed and the problem of determining the optimum variation in the characteristic impedance of a section of long line matching two lines with different characteristic impedances.

The coefficient of reflection R from a section of long line with characteristic impedance $W(z)$ is

$$R = \frac{1}{2} \int_0^1 \frac{d \ln W}{d \xi} e^{-2i h L \xi} d \xi, \quad (1)$$

where L is the length of the section and $\xi = z/L$.

On the other hand, if in the system of equations (12) of Ref. 1 or, what is the same,

Eqs. (39)-(40) of Ref. 3 we neglect (except for the exponent) the difference of wave numbers h_1 and h_2 of the H₀₁ and H₀₂ waves from the free-space wave number (in Section 2 the validity of this approximation is shown by an example), then the amplitude of the forward H₀₂ wave will be

$$D(\sigma) = C \int_0^1 \frac{d \ln a}{d \xi} e^{-i\sigma \xi} d\xi. \quad (2)$$

The relative power losses are equal to $|D|^2$. Here the parameter σ is the phase excursion difference of both waves over the entire length of the transition.

$$\sigma = \int_0^1 (h_1 - h_2) dz; \quad (3)$$

ξ is a new variable, defined as

$$\xi = \frac{1}{\sigma} (\gamma_1 - \gamma_2), \quad (4)$$

where $\gamma_p = \int_0^1 h_p dz$; C is a constant equal to

$$C = -\frac{2\mu_1\mu_p}{\mu_p^2 - \mu_1^2}; \quad (5)$$

μ_p is the p -th root of the equation $J_1(\mu) = 0$. With the same degree of accuracy as exists in Eq. (2), the relation between ξ and z may be written in the form

$$z(\xi) = \frac{L}{a_{av}^2} \int_0^1 a^2 d\xi, \quad (6)$$

where

$$a_{av}^2 = \int_0^1 a^2(\xi) d\xi; \quad (7)$$

a — is the transition radius considered as a function of variable ξ .

Comparing Eqs. (1) and (2), we see that in Eq. (2) D and a are related in the same manner as R and W in (1). The parameter $2kL$ is replaced in Eq. (2) by the parameter σ , which is defined by

$$\sigma = \frac{\mu_p^2 - \mu_1^2}{4\pi} \frac{\lambda L}{a_{av}^2}. \quad (8)$$

σ decreases as frequency increases. Further, the amplitudes of forward parasitic waves created at the transition will increase, since the difference in phase constants of the various modes decreases, as does the mutual compensation of parasitic waves which arise in the various waveguide sections. We will note that in Eq. (1) the wavelength differs from that in Eq. (2), for the expression for the amplitude of the reflected wave includes the sum of the phases of both waves rather than their difference. The functions $a(\xi)$ will be chosen in the same manner as were $W(\xi)$ in the theory of long lines. Given $a(\xi)$, from Eqs. (6) and (7) we can find $z(\xi)$, that is, we can determine the shape of the generatrix in

parametric form. The amplitude of the parasitic wave is determined from Eqs. (8) and (2). The functions given in Ref. 6 will be used in choosing $a(\xi)$. Let it be noted that determinations of the true optimum [7] for the integral of Eq. (2) are of no significance in our case, for Eq. (2) itself, for the problem under discussion, is approximate.

The form of the function $a(\xi)$ depends on the maximum permissible amplitude of the parasitic wave. Assume, for example, that $|D|$ must be of the order of $(3-5) \cdot 10^{-2}$ (the amplitude of the H_{01} wave is assumed to be unity). In this case we may use the function $W(\xi)$ defined by Eq. (8) in Ref. 6. Let us assume further that with $\xi = 0$ and $\xi = 1$ the waveguide radius assumes the assigned values of $a(0)$ and $a(L)$. Then

$$a(\xi) = a(0) e^{\ln q \left(\xi - \frac{0.636}{2\pi} \sin 2\pi \xi \right)}, \quad (9)$$

where $q = a(L)/a(0)$. At the ends of the generatrix defined by this function there are breaks equal to

$$\frac{da}{dz} = 0.364 \ln q \frac{a_{cp}^2}{L} \frac{1}{a}. \quad (10)$$

At the wide end the break is smaller than at the narrow end.

For the variation defined by Eq. (9) the function $|D(\sigma)|$ has the form

$$|D(\sigma)| = \left| C \ln q \frac{\sin \frac{\sigma}{2}}{\frac{\sigma}{2}} \frac{4\pi^2 - 0.364 \sigma^2}{4\pi^2 - \sigma^2} \right|. \quad (11)$$

The plot of $|D(\sigma)| / |\ln q|$ is shown in Fig. 1. With $\sigma > \sigma_{\min} = 3\pi$, $|D(\sigma)| < 0.048 |\ln q|$. Knowing q and denoting the minimum wavelength within the band as λ_{\min} , from Eqs. (7) and (8) we may determine the minimum length of the transition. Let us note that a transition providing sufficiently low loss for the H_{02} wave will provide [within the limits of Eq. (1)] even lower loss for the H_{03} wave, etc.

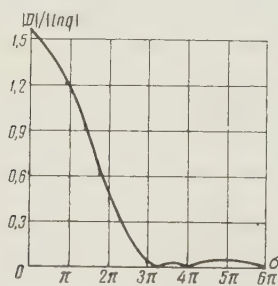


Fig. 1

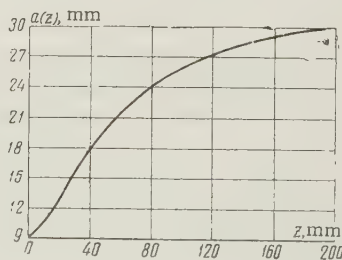


Fig. 2

Let us consider an example. If $a(0) = 9$ mm, $a(L) = 30$ mm and $\lambda_{\min} = 6.6$ mm, then the transition length will be 193.7 mm. At all $\lambda > \lambda_{\min}$ [within the limits of applicability of Eq. (2)], we will find $|D(\sigma)| < 0.06$. For a taper such losses would be permissible at lengths 8 - 9 times greater. The plot of the function $a(z)$ is shown in Fig. 2. The breaks at the ends are equal to 5.4 and 1.6° . In order to achieve a transition with minimum loss it is necessary to use another function from Ref. 6 or other works on the theory of long lines.

2. NUMERICAL METHOD OF DETERMINING THE TRANSFORMATION COEFFICIENT IN THE ABSENCE OF A CRITICAL CROSS-SECTION

Let us now calculate the transformation coefficient without admitting the simplifications introduced in Eq. (2). In calculating the amplitude of the forward wave $|D|$ with the chosen generatrix it is convenient to use the system of equations obtained from Eqs. (39)–(40) of Ref. 3 by converting to a new independent variable ξ and new functions $b(\xi)$ and $d(\xi)$:

$$b(\xi) = B e^{i\gamma_2}, \quad d(\xi) = D e^{-i\gamma_2}. \quad (12)$$

The quantities b and d represent the amplitudes of the forward and backward parasitic waves. The amplitude of the H_{01} wave is considered to be equal to unity over the entire transition.

Differentiation with respect to ξ will be indicated by a dot. The system of equations then takes the form

$$\begin{aligned} \dot{d} + i h_2 d \frac{a^2 L}{a_{av}^2} - \frac{\dot{h}_2}{2 h_2} b &= \frac{C \dot{a}}{a} \frac{h_2 + h_1}{2 \sqrt{h_1 h_2}} e^{-i\gamma_1}, \\ \dot{b} - i h_2 b \frac{a^2 L}{a_{av}^2} - \frac{\dot{h}_2}{2 h_2} d &= \frac{C \dot{a}}{a} \frac{h_2 - h_1}{2 \sqrt{h_1 h_2}} e^{-i\gamma_1}. \end{aligned} \quad (13a)$$

The boundary conditions for this system have the form

$$d(0) = 0, \quad b(1) = 0. \quad (13b)$$

Direct solution of the problem in Eqs. (13) leads to a number of difficulties, for Eq. (13b) includes conditions at different ends of the interval. Equation (13a) is a linear system and hence may be reduced to two Cauchy problems [8]. In the first problem the system of equations coincides with Eq. (13a) and in the second it differs from Eq. (13a) by the absence of the right-hand members. The initial conditions of both problems are

$$d(1) = 1, \quad b(1) = 0. \quad (14)$$

The solution of the system of Eqs. (13) is given in terms of the solutions $d(1)$, $b(1)$ and $d(2)$, $b(2)$ of these Cauchy problems by

$$d(\xi) = d^{(1)}(\xi) - \frac{d^{(1)}(0)}{d^{(2)}(0)} d^{(2)}(\xi), \quad (15a)$$

$$b(\xi) = b^{(1)}(\xi) - \frac{d^{(1)}(0)}{d^{(2)}(0)} b^{(2)}(\xi). \quad (15b)$$

In accordance with Eq. (15) the amplitude of the forward parasitic wave $|D| = |d(1)|$ is

$$|D| = \left| 1 - \frac{d^{(1)}(0)}{d^{(2)}(0)} \right|. \quad (16)$$

Upon separation of the real and imaginary members, the given systems of equations lead to systems of four equations which have been solved by us on the BESM [high-speed electronic computer] by the Runge-Kutta method. Figure 3 gives the results of a calculation for the transition whose generatrix is shown in Fig. 2. The solid line is plotted from the results obtained on the BESM by use of Eqs. (15) and (16) and the broken line is plotted from results of calculation with Eq. (11). Comparison of the curves shows that the assumption made in Section 1 is valid. This confirms the appropriateness of the method chosen for selecting the shape of the generatrix. Over a wide range (even at the wavelength at which the critical cross-section arises) the losses will be but a few decibels higher than in

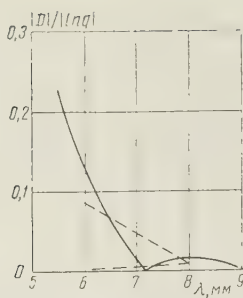


Fig. 3

Eq. (11). However, it will not reflect in detail the shape of the curve for the dependence of $|D|$ on λ in Eq. (11).

The dash-dot curve is plotted from the amplitude of the backward parasitic wave $|b(0)|$ calculated from Eq. (15b). As is to be expected from general considerations, the backward wave under these conditions is extremely small.

3. NUMERICAL METHOD OF DETERMINING THE TRANSFORMATION COEFFICIENT IN THE PRESENCE OF THE CRITICAL CROSS-SECTION

Let us assume at a certain cross-section $z = \tilde{z}$ that the wave number of the H₀₂ wave vanishes: $h_2(\tilde{z}) = 0$. The method of direct numerical calculation from Eq. (13), as used in the preceding section, is not applicable here, for the coefficients in (13a) have singularities at $z = \tilde{z}$. For the sake of definiteness let us discuss a tapered transition. The conversion losses in the transition are determined by the amplitude of the parasitic wave reflected into the wide waveguide, $|B(0)|$. From Eqs. (11) - (13) of Ref. 4 for $B(0)$ we may arrive at the expression

$$B(0) = \sqrt{\frac{h_1(0)}{h_2(0)}} \left\{ e^{-2i(\tilde{\gamma}_2 + \frac{\pi}{4})} \int_0^{z_L} C \frac{h_2 + h_1}{2\sqrt{h_1 h_2}} \frac{d \ln a}{dz} e^{-i(\gamma_1 - \gamma_2)} dz + \right. \\ \left. + \int_0^{z_L} C \frac{h_2 - h_1}{2\sqrt{h_1 h_2}} \frac{d \ln a}{dz} e^{-i(\gamma_1 + \gamma_2)} dz - e^{-i(\tilde{\gamma}_1 + \frac{\pi}{4})} \int_{z_L}^L C \frac{e^{-i\gamma_1}}{\sqrt{h_1}} \frac{d \ln a}{dz} (V' + i h_1 V) dz \right\}. \quad (17)$$

Here z_L is any point for which $z_L < \tilde{z}$. Let us choose z_L extremely close to the critical cross-section. The functions V and dV/dz are expressed as Airy functions (see Refs. 3 and 4). The evaluation of the first two integrals in Eq. (17) will reduce to the solution of a system of differential equations. Let us denote these integrals by J_1 and J_2 , respectively, and introduce the functions

$$u = J_1 e^{-i\gamma_1}, \quad v = J_2 e^{i\gamma_2}. \quad (18)$$

These functions satisfy an inhomogeneous system of equations in the variable ξ differing from (13a) only in the absence of terms containing h_2 . The initial conditions of the system have the form

$$u(0) = 0, \quad v(0) = 0. \quad (19)$$

As in system (13), we calculated the first two components in Eq. (17) by the Runge-Kutta method.

The integrand in the third component of Eq. (17) decreases rapidly in the region $z > \tilde{z}$; hence the third integral was calculated by the Gaussian quadrature formula with eight abscissas.

As an example we calculated a series of generatrices for various values of $a(0)$ and $a(L)$. The curves for these are given in Fig. 4 (Curves I - IV). The table lists the values of $|B(0)|$ obtained for these transitions by the method described. The calculations were performed for $\lambda = 8$ mm.

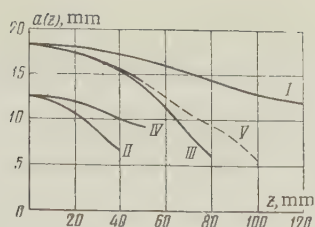


Figure 4

| Curve No. | $ B(0) $ |
|-----------|----------------------|
| I | $1.7 \cdot 10^{-2}$ |
| II | $0.56 \cdot 10^{-2}$ |
| III | $26 \cdot 10^{-2}$ |
| IV | $1.2 \cdot 10^{-2}$ |

As is seen from the table, in certain cases the form of the generatrix as obtained by the method described in Section 1 may prove unsatisfactory in the presence of the critical cross-section.

Analysis shows that the first integral in Equation (17) has greater effect on $|B(0)|$ than other components. Hence it is sometimes necessary to choose $a(z)$ so that this integral and not integral (2) is as small as possible.

For this purpose let us introduce the function $F(a)$, defined by the condition

$$dF = \frac{da}{a} \frac{h_1 + h_2}{2 \sqrt{h_1 h_2}} \quad (20)$$

$F(a)$ may be determined easily from Equation (20), but due to its unwieldiness we shall not write it out. Let us further choose $a(\xi)$ so that $F(a)$ is proportional to the function chosen [optimum for Eq. (1)] in $W(\xi)$. Here the variable ξ is determined as in Eqs. (3) and (4) by replacing L with z_L . We will note that the expression for $F(a)$ includes frequency (wave numbers h_1 and h_2 depend on k) and we choose the most effective frequency of the band. For other frequencies of the band a transition of modified form will not, generally speaking, be better than the transition obtained in Section 1.

Having determined $F(a)$ and $F(\xi)$, let us determine the dependence of a on ξ . The dependence of z on ξ is determined from the condition

$$\frac{dz}{z_L} = \frac{h_1 + h_2}{2k} \frac{a^2(\xi) d\xi}{\int_0^1 \frac{h_1 + h_2}{2k} a^2(\xi) d\xi} \quad (21)$$

Knowing $a(z)$, we can, as in the preceding case, find the length z_L . The form of the generatrix for $z > z_L$ affects only the value of the third component in Eq. (17) and the quantity $|B(0)|$ for wavelengths shorter than that for which the optimum $a(z)^*$ (for our purposes) was determined.

In Fig. 4 Curves III-V are given for the same conditions, while Curve V is derived by the method described. As seen from the figure, with the exception of the region adjacent to the critical cross-section, for the wavelength chosen ($\lambda = 8$ mm) it almost completely repeats curve III. Hence, a transition with this form of generatrix yields for other wavelengths approximately the same result as a transition with generatrix III. The value of $|B(0)|$ obtained for curve V was 3.4%.

It is obvious that the method proposed here, beginning with determination of $F(a)$ from Eq. (20), may also be used for designing transitions in the absence of a critical cross-section. However, satisfactory results may also be obtained by the simpler method proposed in Section 1.

CONCLUSION

The calculation of losses for a chosen form of $a(z)$ in the absence of a critical cross-section for each value of λ requires approximately 5 minutes on the BESM, while in the presence of a critical cross-section it requires 10 minutes. The design of wideband

* In practical calculations we chose z_L so that the condition $|t| \ll N^{1/2}$, $|t| \gg 1$, was fulfilled, where t and N are determined from Ref. 3.

waveguide transitions with extremely low conversion loss (on the order of 40-50 db) and relatively short length probably requires the testing of several forms of $a(z)$ and perhaps also consideration of the change in amplitude of the H_{01} wave along the transition, and hence will require several hours of machine time.

We take this opportunity to express our thanks to co-workers of the theoretical-design group of IRE AN SSSR for their assistance in the calculations.

REFERENCES

1. B. Z. Katsenelenbaum, Nonuniform waveguides with slowly changing parameters, Dokl. AN SSSR, 1955, 102, 4, 711.
2. H. Unger, Circular waveguide taper of improved design, Bell System Techn. J., 1958, 37, 4, 899.
3. B. Z. Katsenelenbaum, Long symmetrical waveguide transition for H_{01} waves, Radiotekhnika i elektronika, 1957, 2, 5, 531.
4. B. Z. Katsenelenbaum, Critical cross-sections in nonuniform waveguides, Dokl. AN SSSR, 1958, 123, 1, 53.
5. B. Z. Katsenelenbaum, On the theory of nonuniform waveguides with slowly changing parameters, Congres International Circuits et Antennes Hyperfréquences, Paris, 1957, part I, p. 124.
6. J. Willis, N. Sihna, Nonuniform transmission lines as impedance transformers, Proc. IEE, 1956, 103B, 8, 166.
7. R. Klopfenstein, A transmission line taper of improved design, Proc. IRE, 1956, 44, 1, 31.
8. L. Kollats, Numerical Methods of Solving Differential Equations, IL, 1953.

Institute of Radio Engineering
and Electronics, AN SSSR

Submitted to the Editors
24 June 1960

SYNTHESIS OF STEPPED DIRECTIONAL COUPLERS

A. L. Fel'dshteyn

The paper discusses the theory of directional couplers in the form of a chain of sections of coupled lines. The system possesses theoretically ideal directivity and matching within an infinite bandwidth and optimum (Chebyshev or maximally linear) frequency response in the division of power between the branches. Also discussed are several properties of a system having an infinite number of steps. Several laborious design calculations are tabulated.

INTRODUCTION

A section of uniform coupled lines may serve as a directional coupler within a wide range of frequencies [1, 2]. In a cascade connection of several such sections the bandwidth of the system is increased [3]. It is of interest to examine the theory of the synthesis of optimum cascade connection of coupled lines. Criteria of optimality may differ; hence we shall discuss two cases: a directional coupler with minimum number of sections (Chebyshev or iso-extremal characteristic) and a coupler with maximally a linear characteristic. The physically attainable properties of directional systems are also discussed, since they are intimately related to the synthesis problem.

Below we present the principal assumptions of this theory and several of its applications.

1. SINGLE-STEP DIRECTIONAL COUPLER

Let us examine a uniform symmetrically coupled line (Figure 1) the equations of which, according to Reference 1, have the form*

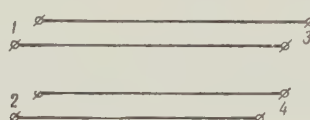


Figure 1

$$\begin{aligned} U_1 &= U_3 \cos ml + j(\rho I_3 + r I_4) \sin ml, \\ U_2 &= U_4 \cos ml + j(r I_3 + \rho I_4) \sin ml, \\ I_1 &= I_3 \cos ml + j\left(\frac{U_3}{w} - \frac{U_4}{v}\right) \sin ml, \\ I_2 &= I_4 \cos ml + j\left(\frac{U_4}{w} - \frac{U_3}{v}\right) \sin ml, \end{aligned} \quad (1)$$

where ρ and w are the characteristic impedances and r and v are the current and voltage coupling impedances respectively.

For such a line, as is known [1], the following relationships hold

$$w = \frac{\rho^2 - r^2}{\rho}, \quad v = \frac{\rho^2 - r^2}{r}, \quad rv = \rho w. \quad (2)$$

Let us assume that the load resistances at all outputs of the eight-terminal network are identical, $R = 1$ ohm, and let us introduce the condition of ideal directivity and matching of the system

$$rv = R^2 = 1. \quad (3)$$

Under these conditions the wave transmission matrix of a single-step coupled line has the form

$$[T] = \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} = \begin{bmatrix} \cos \Theta + j\rho \sin \Theta & 0 & 0 & -jr \sin \Theta \\ 0 & \cos \Theta + j\rho \sin \Theta & -jr \sin \Theta & 0 \\ 0 & jr \sin \Theta & \cos \Theta - j\rho \sin \Theta & 0 \\ jr \sin \Theta & 0 & 0 & \cos \Theta - j\rho \sin \Theta \end{bmatrix} \quad (4)$$

where r and ρ are normalized ($r \sim r/R$ and $\rho \sim \rho/R$). From Equation (4) it is easy to find the desired scattering matrix elements of one step:

$$S_{11} = 0, \quad S_{14} = 0, \quad (5)$$

*The notation is changed slightly.

$$S_{13} = \frac{1}{\cos \Theta + j\rho \sin \Theta}, \quad S_{12} = \frac{j\rho \sin \Theta}{\cos \Theta + j\rho \sin \Theta}. \quad (6)$$

2. MULTI-STEP DIRECTIONAL COUPLER

A cascade connection of any number of uniform coupled lines with balanced coupling ($rv = 1$), discussed in the preceding section, will be referred to as a multi-step directional coupler. Such a system is ideally directive and matched over an infinite bandwidth. Thus, confronted with the problem of the synthesis of an optimum multi-step directional coupler, the only parameter requiring optimization is the power division factor between branches 2 and 3 of the system:

$$\eta = \frac{|S_{12}|^2}{|S_{13}|^2}. \quad (7)$$

In the synthesis of this frequency function it is necessary to keep in mind the following.

1. The power division factor is numerically equal to the square of the absolute value of element $M(T_{41})$ of the wave transmission matrix $[T]$ of the system*.

2. The power division factor is a polynomial in $\cos^2 \Theta$ **:

$$\eta = \alpha_0 + \alpha_2 \cos^2 \Theta + \alpha_4 \cos^4 \Theta + \dots + \alpha_{2n} \cos^{2n} \Theta, \quad (8)$$

where

$$\Theta = ml = 2\pi \frac{l}{\lambda}; \quad (9)$$

is the length of the step; α_i are real constants.

Due to the unitarity of the matrix $[S]$

$$|S_{12}|^2 + |S_{13}|^2 = 1, \quad (10)$$

and, consequently, the quantity η fully defines the system. Let it be required that the function $\eta(\Theta)$ have the form

$$\eta = \beta^2 - h^2 T_n^2 \left(\frac{\cos \Theta}{S} \right), \quad (11)$$

where β is a constant; h and S are amplitude and scale factors;

$$T_n(x) = \cos n \arccos x \quad (12)$$

is a Chebyshev polynomial of the first kind and of the n th order. The approximate shape of the frequency characteristic corresponding to Eq. (11) is shown in Figure 2.

The relation between the parameters of a multi-step system follows from an evident condition: with $\Theta = 0$ $\eta = 0$. Then from Equations (11) and (12) we obtain an expression defining the attainable passband of a multi-step system:

$$S = \frac{1}{\operatorname{ch} \frac{1}{n} \operatorname{arccch} \frac{\beta}{h}}. \quad (13)$$

Thus, the bandwidth of constant power division (the "passband") in a directional coupler is fully determined by the number of sections n , the nominal level of power division β^2 and the allowable deviation from the given level h^2 .

*The matrix elements $[T]$ of the system are designated by italic capitals in the same sequence as given in Equation (4) for a single step.

**See Appendix 1.

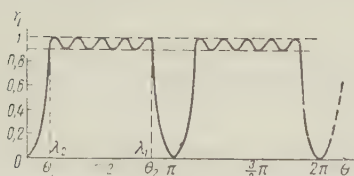


Figure 2

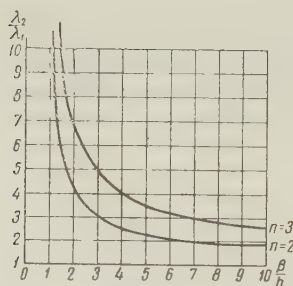


Figure 3

After the scale coefficient S has been determined from Equation (13), the band overlap factor χ is found from the equation

$$\chi = \frac{\lambda_2}{\lambda_1} = \frac{\pi - \arccos S}{\arccos S}, \quad (14)$$

where λ_2 and λ_1 are the cutoff wavelengths of the system with a given tolerance. Equations (13) and (14) are illustrated by the plots in Figure 3 for $n = 2$ and 3. These curves bound the physically attainable bandwidth of a directional coupler with given parameters n , β and h .

From analysis of Equation (11) it follows that the length of a single step in the system is defined by

$$l = \frac{1}{2} \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}. \quad (15)$$

In a number of cases it is of interest to consider a number of cases it is of interest to consider a multi-step directional coupler with a maximally linear power division frequency characteristic (Figure 4). Let us express this characteristic in the form

$$\eta = \beta^2 - \beta^2 \cos^{2n} \Theta \quad (16)$$

(a special case of this relationship - the simplest single-step ($n = 1$) system - was discussed in Section 1).

Calculation of the coupling impedances of the coupler steps in both of the types mentioned may be performed by the method of undetermined coefficients, as was done in References 4 and 5 for stepped transitions. The design formulas are derived by comparing the coefficients of the same powers of $\sin^2 \Theta$ in Equation (11) or (16) and in the expression for this function as determined by direct multiplication of the matrices $[T]$ of the steps of the directional coupler. The explanation is given in Appendix 2. This same appendix contains tables of values of r calculated by this method for the case of a two-step ideal directional coupler with a Chebyshev power division characteristic.

3. LIMITING MULTI-STEP SYSTEM

It was shown above that a multi-step directional coupler has a periodic frequency characteristic (Figures 2 and 4) and, consequently, that the passband of such a coupler is always limited at the low- and high-frequency ends. There may arise cases in which a constant

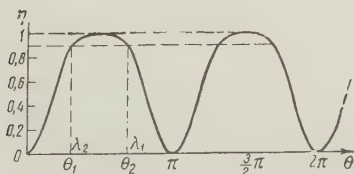


Figure 4

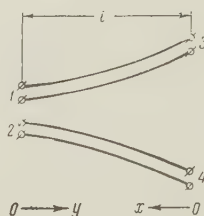


Figure 5

transition loss within the coupler is required within an unlimited range of frequencies. An excellent approximation to such a characteristic is obtained in the case of the limiting multi-step system, in other words, in the case of coupled tapered lines. Let us use the limiting relation of the first approximation for the power-division constant of loosely-coupled non-uniform lines [6]

$$\eta = S_{12} = jm \int_0^l r(y) e^{-j2my} dy, \quad (17)$$

where $r(y)$ is the coupling distribution function; $y = l - x$ is a coordinate read from the origin (Figure 5). Assuming that the coupling changes exponentially

$$r(y) = r_0 e^{-\alpha y}$$

and integrating in Equation (17), we have

$$\eta = r_0 \frac{m^2}{\alpha^2 + (2m)^2} \{1 + e^{-2\alpha l} - 2e^{-\alpha l} \cos 2ml\}. \quad (18)$$

A typical characteristic plotted from this equation is shown in Figure 6. The broken line represents the limit curve (with $l \rightarrow \infty$). It is evident from the curves that the passband of a directional coupler in the form of tapered lines is limited only at low frequencies. At this point we may compare stepped transitions and tapered transitions: while the first have a periodic characteristic and their passband is limited on both sides, the passband of the latter is limited only on the low-frequency side. The comparison may be extended as follows: with a given finite passband a stepped system will always be shorter

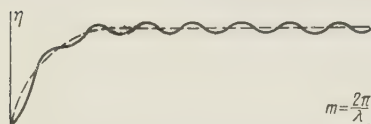


Figure 6

than a tapered system, as is the case with ordinary nonuniform lines [6].

CONCLUSIONS

Modern synthesis of circuits is based on frequency characteristics in the form of a Chebyshev polynomial, Zolotarev fractions and the so-called maximally linear function. In recent years this method has found wide application in microwave techniques. Particular interest attaches to the case where synthesis is applied directly, without equivalent circuits — by an appropriate substitution for a frequency variable ("frequency conversion"). In this respect, devices consisting of uniform line sections of identical length are extremely convenient and are successfully applied in the synthesis of matching transformers and filters [7, 8].

In the present paper this method was extended to directional couplers. These devices, possessing (with a proper choice of parameters) ideal directivity and matching over an unlimited range of frequencies, have a polynomial frequency characteristic of power division between branches. This polynomial relationship permits evolving a synthesis for directional couplers with Chebyshev and maximally-linear characteristics which in many respects is similar to the synthesis of stepped transitions and stepped filters.

APPENDIX 1

It follows from Reference 9 that

$$\frac{S_{12}}{S_{13}} = \frac{BI - AM}{F}, \quad (I)$$

but it is evident from Equation (4) that for our case

$$I = 0, \quad A = F, \quad (II)$$

consequently,

$$\frac{S_{12}}{S_{13}} = -M, \quad (III)$$

and

$$|S_{12}|^2 + |S_{13}|^2 = 1.$$

Thus, the element M of the wave transmission matrix wholly defines the power division between branches 2 and 3. Let us now investigate the nature of the frequency dependence of this element. It is evident from Equation (4) that in multiplying any number of such matrices each element of the resulting matrix is the sum of the products of the following terms:

$$j\rho_q \sin \theta, \quad \cos \theta + j\rho_i \sin \theta, \quad (IV)$$

that is, the possible combinations

$$\sin^n \theta, \quad \cos^m \theta, \quad \sin^n \theta \cos^m \theta. \quad (V)$$

We are interested in the modulus of Equation (III); if

$$\frac{S_{12}}{S_{13}} = x + jy,$$

then

$$\eta = \left| \frac{S_{12}}{S_{13}} \right|^2 = x^2 + y^2,$$

that is, terms of type (V) acquire the form of even functions

$$\sin^{2n} \theta, \quad \cos^{2m} \theta, \quad \sin^{2n} \theta \cos^{2m} \theta. \quad (VI)$$

All even powers of sines and cosines may, after elementary conversion, be expressed in terms of $\cos^2 \theta$ or $\sin^2 \theta$. Thus, the power division constant is a polynomial in $\cos^2 \theta$:

$$\eta = \alpha_0 + \alpha_2 \cos^2 \theta + \alpha_4 \cos^4 \theta + \dots + \alpha_{2n} \cos^{2n} \theta. \quad (VII)$$

APPENDIX 2

Multiplying the matrices of two steps, we find

$$M = \alpha \sin^2 \theta + jk \sin \theta \cos \theta, \quad (VIII)$$

where the following designations are used:

$$\alpha = \rho_1 r_2 - \rho_2 r_1; \quad (IX)$$

$$k = r_1 + r_2. \quad (X)$$

Hence,

$$\eta = |M|^2 = \sin^2 \theta \{k^2 - (k^2 - \alpha^2) \sin^2 \theta\}. \quad (XI)$$

On the other hand, it is necessary that

$$\eta = \beta^2 - h^2 T_n^2 \left(\frac{\cos \theta}{S} \right) = \sin^2 \theta \left\{ 4h^2 \left(\frac{2}{S^4} - \frac{1}{S^2} \right) - 4 \frac{h^2}{S^4} \sin^2 \theta \right\}. \quad (XII)$$

Comparing the coefficients of the same powers of $\sin^2 \theta$ in Equations (XI) and (XII), we obtain the system of equations

$$\begin{cases} (r_1 + r_2)^2 = 4h^2 \left(\frac{2}{S^4} - \frac{1}{S^2} \right), \\ (r_1 + r_2)^2 - (r_2 \rho_1 - r_1 \rho_2)^2 = 4 \frac{h^2}{S^4}. \end{cases} \quad \begin{matrix} (XIII) \\ (XIV) \end{matrix}$$

Considering that

$$\rho_1 = \sqrt{1 + r_1^2}, \quad \rho_2 = \sqrt{1 + r_2^2}, \quad (XV)$$

we obtain after conversion

$$A (r_1 r_2)^2 + 2B (r_1 r_2) - B^2 = 0, \quad (XVI)$$

where

$$A = 4h^2 \left(\frac{1}{S^4} - \frac{1}{S^2} \right); \quad (XVII)$$

$$B = \frac{2h^2}{S^4}. \quad (XVIII)$$

Solution of Equation (XVI) yields

$$r_1 = \frac{1}{2} \left\{ k + \sqrt{k^2 + 4C \left(1 \pm \sqrt{1 + A} \right)} \right\}, \quad (XIX)$$

$$r_2 = \frac{1}{2} \left\{ k - \sqrt{k^2 + 4C \left(1 - \frac{1}{\sqrt{1+A}} \right)} \right\}, \quad (\text{XX})$$

where

$$C = \frac{B}{A} = \frac{1}{2(S^2 - 1)}.$$

| Allowance for transition loss δ , db | Coupling impedance | | Overlap coefficient λ_2/λ_1 |
|---|--------------------|--------|---|
| | r_1 | r_2 | |
| 0.1 | 1.3290 | 0.2311 | 2.3461 |
| 0.2 | 1.3515 | 0.2609 | 2.7790 |
| 0.3 | 1.3677 | 0.2911 | 3.2137 |
| 0.5 | 1.3796 | 0.3235 | 3.6897 |
| 1.0 | 1.3913 | 0.3970 | 4.8500 |

We present a table of values of r_1 and r_2 for a two-step Chebyshev directional coupler with equal power division between branches 2 and 3 ($\eta_0 = 1$, transition loss $L_0 = 3$ db).

REFERENCES

1. A.A. Pistol'kors, Antennas, Svyaz'izdat, 1947.
2. A.R. Vol'pert, Theoretical and experimental investigation of the transmission-line reflectometer, Radiotekhnika, 1941, 2, 2, 313.
3. J.K. Shimzu, Strip-line 3-db directional couplers, IRE Wescon Convention Record, 1957, August, Part I, 4.
4. R. Kollen, Theory and calculation of wideband multi-sectional quarter-wave transformers, Sb. Vopr. radiolokatsionnoy tekhniki (Radar engineering symposium), IL, 1955, 5(29).
5. A.L. Fel'dshteyn, L.R. Yavich, Engineering calculation of Chebyshev stepped transitions, Radiotekhnika, 1960, 15, 1, 3.
6. A.L. Fel'dshteyn, L.R. Yavich, Comparison of stepped and tapered transitions, Radiotekhnika i elektronika, 1959, 4, 3, 527.
7. N. Rible, General synthesis of quarter-wave impedance transformers, Sb. Vopr. radiolokatsionnoy tekhniki, IL, 1957, 4(40), 3.
8. A.L. Fel'dshteyn, Synthesis of stepped filters, Voprosy radioelektroniki (Problems in radio electronics), series XII, 1959, 19, 40.
9. B.B. Lagovier, Calculation of complex waveguide systems by means of the theory of eight-terminal networks, Sb. Elektronika (Electronics symposium), 1958, 6, 64.

Submitted to the editors 9 June 1960

ATTENUATION MEASUREMENTS IN WAVEGUIDES

Yu. N. Kazantsev

This paper proposes a method for the measurement of attenuation in waveguide elements by determining the standing-wave ratio in the waveguide in front of a cavity formed by the piece to be measured, a diaphragm and a shorting plunger, at resonance. Measurement setups are presented for the H_{01} wave in circular waveguides and formulas are given for error determina

1. MEASUREMENT PROCEDURE

The resonant cavity method is usually used in measuring attenuation in relatively short waveguide sections. This method consists essentially of the measurement of the Q of the cavity formed by the waveguide section under investigation and two metal end plates. The Q is then recalculated taking into account the effect of losses in the end portions on the attenuation in the section.

Methods of Q determination at microwave frequencies may be divided into two classes: (a) determining the Q from the duration of free oscillations; (b) determining the Q from the width of the resonance curve.

The method of determining the Q from the duration of free oscillations, described in Reference 1, is suitable for systems with extremely high Q ; with a decrease in Q the measurement error increases considerably.

Methods of determining Q from the width of the resonance curve find their widest application in microwave techniques [2, 3]. However, in a number of cases they prove unsuitable owing to the high stability requirements for the signal generator.

The method described below permits the measurement of an extremely wide range of attenuation values with the use of simple equipment. In addition, the method does not place high stability requirements on the signal generator. Attenuation within the section is determined from the relative power absorbed in the cavity at resonance.

Let us determine the relationship between the value of this power and the parameters of the cavity resonator. For this purpose we shall examine the cavity formed by diaphragm 1 and a shorting plunger 2 placed within a waveguide 3 which is matched with the signal generator (Figure 1).

The diaphragm must pass the given wave mode without exciting other, unwanted modes.

Under conditions near resonance the field within the cavity may be regarded as the sum of two opposed traveling waves of equal amplitude. The expression for the amplitude of these waves was derived in Reference 4 by the method of superposition of waves multi-reflected from the end faces. At the resonant frequency, after transformation, this expression takes the form

$$u = \frac{2A}{\eta_{dp}}, \quad (1)$$

where A is the amplitude of the traveling-wave field to the right of the diaphragm in the case where, instead of a shorting plunger, there is connected a matching load; $\eta_{dp} \ll 1$ is a quantity indicating that part of the power which is lost by the wave within the cavity (the quantity η_{dp} includes both the thermal losses and radiation losses within the waveguide).

Before proceeding to the equation defining the power absorbed in the cavity, let us explain the physical meaning of certain quantities. If we designate the power of the incident wave as P_1 and the transfer constant of the diaphragm was η_1 , then the product $P_1\eta_1$ represents the power of the wave passing through the diaphragm in traveling-wave operation (diaphragm within a waveguide matched with the generator and load). The ratio u/A indicates the number of times by which the wave amplitude within the cavity has increased as the result of multiple reflections from the end faces; the difference $\eta_{dp} - \eta_1$ defines the thermal losses in the cavity.

Now it is easily seen that the power P absorbed in the cavity is defined by

$$P = P_1\eta_1 \left(\frac{u}{A} \right)^2 (\eta_{dp} - \eta_1). \quad (2)$$

Substituting Equation (1) in Equation (2), we obtain

$$P = 4P_1\eta_1 \frac{\eta_{dp} - \eta_1}{\eta_{dp}^2}. \quad (3)$$

Assuming in Equation (3) that $P=P_1$, let us determine the condition for the total absorption of the power of the wave reaching the cavity from the waveguide:

$$\frac{\eta_1}{\eta_{dp}} = 0.5. \quad (4)$$

From Equation (3) we can also easily express the thermal loss



Figure 1. Diaphragm and plunger in waveguide.

$$\eta = \eta_{dp} - \eta_1 \quad (5)$$

in terms of the diaphragm transfer constant and the relative power $p = P/P_1$ absorbed in the cavity:

$$\eta = \eta_1 \left[\left(\frac{2}{p} - 1 \right) \pm \frac{2}{p} \sqrt{1-p} \right]. \quad (6)$$

Equation (6) can be written in simpler form by expressing the quantity p in terms of the standing-wave ratio in the waveguide in front of the cavity. Since $p = 1 - k^2$ (k is the modulus of the reflection coefficient of the cavity) and $\rho = (1+k) / (1-k)$, then

$$\eta = \eta_1 \rho \quad (7')$$

for $\eta > \eta_1$ [plus sign in Equation (6)] and

$$\eta = \eta_1 / \rho \quad (7'')$$

for $\eta < \eta_1$ [minus sign in Equation (6)].

From Equations (7) it is evident that for the determination of the quantity η it suffices to know the transfer constant of the diaphragm in traveling-wave operation as well as the standing-wave ratio in front of the cavity at resonance. The quantity η includes losses in the side walls of the cavity η_{ct} , losses in the plunger η_{t1} and the diaphragm η_{t2} of the wave within the cavity

$$\eta = \eta_{ct} \pm \eta_{t1} \pm \eta_{t2}. \quad (8)$$

The principal part of the setup for attenuation measurement is the cavity, which consists of the two end faces of the diaphragm and plunger and a waveguide section of one or two half-wavelengths. The cavity portion may include a test piece with length of $n \frac{\lambda_B}{2}$, where n is an integer and λ_B is the waveguide wavelength.

The number η is determined for two cases: (a) the piece is included (η_B) and (b) the piece absent (η_0); then the losses in the piece are defined as the difference of η_B and η_0^* :

$$\eta_{\text{piece}} = \eta_B - \eta_0. \quad (9)$$

From Equations (7) and (9) we can write the following expressions for the attenuation within the piece (in decibels):

$$\begin{aligned} \eta_{\text{piece}} &= 2,17 \eta_1 (\rho - \rho_0) \quad \text{with } \eta_1 < \eta_0 < \eta_B, \\ \eta_{\text{piece}} &= \frac{2,17 \eta_1 (\rho_0 - \rho)}{\rho \rho_0} \quad \text{with } \eta_1 > \eta_B > \eta_0, \\ \eta_{\text{piece}} &= \frac{2,17 \eta_1 (\rho \rho_0 - 1)}{\rho_0} \quad \text{with } \eta_0 < \eta_1 < \eta_B, \end{aligned} \quad (10)$$

where ρ and ρ_0 are the standing-wave ratios in front of the cavity with and without the connection of the piece, respectively. These expressions, as in the case of the initial Equation (1), are valid under the condition that $\eta_{dp} = \eta_{\text{piece}} + \eta_0 + \eta_1 \ll 1$, which is easily satisfied.

From Equations (10) it is easy to derive the equations for calculation of the total relative error in measurement:

$$\begin{aligned} \frac{\Delta \eta_{\text{piece}}}{\eta_{\text{piece}}} &= \frac{\Delta \eta_1}{\eta_1} + \frac{\Delta \rho}{\rho - \rho_0} + \frac{\Delta \rho}{\rho - \rho_0} \quad \text{with } \eta_1 < \eta_0 < \eta_B, \\ \frac{\Delta \eta_{\text{piece}}}{\eta_{\text{piece}}} &= \frac{\Delta \eta_1}{\eta_1} + \frac{\rho_0 \Delta \rho}{\rho (\rho_0 - \rho)} + \frac{\rho \Delta \rho_0}{\rho_0 (\rho_0 - \rho)} \quad \text{with } \eta_1 > \eta_B > \eta_0, \\ \frac{\Delta \eta_{\text{piece}}}{\eta_{\text{piece}}} &= \frac{\Delta \eta_1}{\eta_1} + \frac{\rho_0 \Delta \rho}{\rho \rho_0 - 1} + \frac{\Delta \rho_0}{\rho_0 (\rho \rho_0 - 1)} \quad \text{with } \eta_0 < \eta_1 < \eta_B. \end{aligned} \quad (11)$$

We will note that the case $\eta_0 < \eta_1 < \eta_B$ corresponds to the maximum range of measurable attenuation at one and the same maximum standing-wave ratio.

Below we describe two variants of the setup for attenuation measurement of the H_{01} wave in circular waveguides.

*If the length of the test piece differs considerably from $n \frac{\lambda_B}{2}$, there is introduced into the measurements an additional error, which is extremely small under the condition that $n \gg 1$.

2. MEASUREMENT SETUP

Variant 1. Indication is achieved by means of a directional coupler within a circular waveguide.

The block diagram of the measurement setup is shown in Figure 2a. The signal from klystron generator 1, modulated by a sawtooth voltage, is applied through decoupling attenuators 3 and 3' and calibrated attenuator 2 to an H_{01} wave exciter in circular waveguide 4 [5, 6].

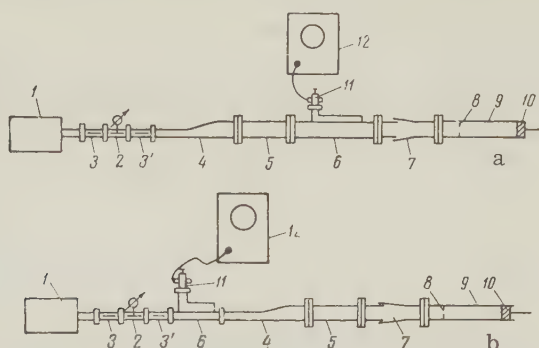


Figure 2. Block diagram of measurement setup.

a) First variant; b) second variant

Passing through helical filter 5, a backward-wave directional coupler in circular waveguide 6 [7, 8] and telescoping junction 7, the H_{01} wave reaches the diaphragm with an aperture system 8.

The diaphragm excites the cavity of test piece 9. Part of the power of the backward wave from detector 11 is applied to oscillograph 12. In tuning the cavity by means of plunger 10 through the klystron range (reproduced on the oscillograph screen) a sharp dip appears (Figure 3a). The standing-wave ratio in the circular waveguide in front of the diaphragm is easily determined by comparing (by means of the calibrated attenuator 2) the height of the response with the level of the dip.*

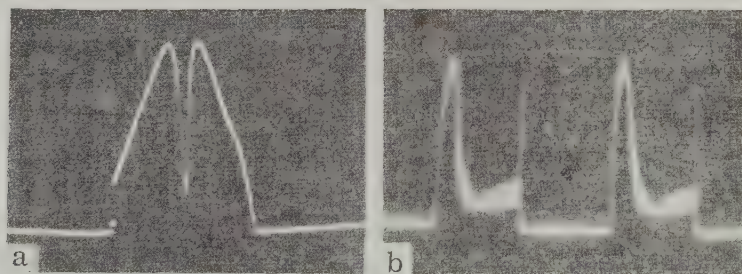


Figure 3. Oscilloscope pattern in tuning.

Variant 2. Indication is achieved by means of a directional coupler in a rectangular waveguide.

The block diagram of the measurement setup for this variant is shown in Figure 2b. In this case the backward-wave directional coupler is placed within the rectangular waveguide before the H_{01} wave exciter.

*In practice the klystron signal is conveniently modulated by rectangular pulses. Then at resonance, due to parasitic frequency modulation, only part of the pulse "collapses" (Figure 3b).

This variant is simpler in construction, for it does not require a directional coupler within a circular waveguide, but the accuracy of measurements, as shown below, is lower than in variant 1.

3. ATTENUATION MEASUREMENTS IN WAVEGUIDE ELEMENTS

The above method has been used in attenuation measurements in straight pieces as well as in curved pieces. Attenuation of the H_{01} wave has been determined in periodic (annular and helical) waveguides with dielectric and metallic sheaths [9, 10]. Attenuation of the H_{01} wave in curved sections of these waveguides has also been investigated [11].

Let us dwell in greater detail on the process of measurement. As has been stated, attenuation within the cavity due to the two kinds of dissipation of the wave is determined from the standing-wave ratio in the waveguide in front of the diaphragm in resonant operation. The standing-wave ratio is measured by comparing the height of the pulse on the oscillograph screen with the level of the dip. Due to the presence of reflections from the H_{01} wave exciter and attenuators, and also due to the fact that the backward-wave coupler has finite directivity, the height of the pulse and of the dip vary upon displacement of the diaphragm and the cavity by means of the telescopic junction.

Let us examine the influence of the above factors on the results of measurement of the standing-wave ratio (SWR) or the coefficient of reflection from the cavity for both variants of the measurement setup. Since the SWR is easily expressed in terms of the reflection coefficient subsequent discussion will deal with the reflection coefficient k .

In the circuit shown in Figure 2a the power of the wave at the detector on the directional coupler will vary with displacement of the diaphragm from

$$\text{to } A' = B(k^2 + \delta_1) \quad (12')$$

$$A'' = B(k^2 - \delta_1), \quad (12'')$$

where B_B is a constant for the given setup, while δ_1 depends on the coefficients of reflection from the cavity and the exciter from the direction of the circular waveguide as well as on the directivity coefficient of the coupler.

The relationships in Equations (12) are valid when

$$k_B \ll 1, \quad k_H \ll 1, \quad (13)$$

where k_B is the modulus of the coefficient of reflection from the exciter and k_H is the power directivity of the coupler*.

If it is considered that in the absence of resonance practically all the power is reflected from the diaphragm and, consequently, the absolute value of the reflection coefficient is equal to unity, then in the case of resonance the absolute value of the corresponding reflection coefficient k is

$$k = \sqrt{\frac{A' + A''}{A'_0 + A''_0}}, \quad (14)$$

where A' and A'' are the maximum and minimum values of the backward-wave power at resonance at the detector in the directional coupler; A'_0 and A''_0 are the corresponding values in the absence of resonance.

From Equation (14) it is seen that in order to determine the reflection coefficient it is sufficient to compare the values of A' and A'' (dip level) and A'_0 and A''_0 (pulse level). As was stated above, this comparison is performed by means of a precisely calibrated attenuator.

In the circuit shown in Figure 2b the coefficient of reflection from the diaphragm within the circular waveguide at resonance is determined by comparison of the reflected waves in the rectangular waveguide up to the H_{01} exciter both at resonance and off resonance.

Upon displacement of the diaphragm, the wave power at the detector in the directional coupler will vary from

*By k_B is meant the absolute value of the total reflection coefficient resulting from the presence of the exciter and from incomplete matching of the decoupling attenuator (3') with the line; k_H takes into account the coefficient of reflection from the power-matching termination of the directional coupler.

to

$$C' = D(k^2 + \delta_2) \quad (15')$$

$$C'' = D(k^2 - \delta_2), \quad (15'')$$

where D is a constant for the given setup; δ_2 depends on the coefficient of reflection from the cavity and the decoupling attenuator as well as on the directivity of the coupler and the characteristics of the H_{01} wave exciter.

Equations (15) are valid when

$$s_{11} \ll s_{12}^2 k, \quad s_{22} \ll 1, \quad k_n \ll s_{1n}^2 k, \quad k_a \ll 1, \quad (16)$$

where s_{11} , s_{22} and s_{12} are the scattering matrix elements of the exciter-filter junction [12]; k_a is the modulus of the coefficient of reflection from the decoupling attenuator.

If it is considered that in the absence of resonance practically all the power is reflected from the diaphragm, then on the basis of Equations (15) we can write the following expression for the modulus of the coefficient of reflection from the cavity at resonance:

$$k = \sqrt{\frac{C' + C''}{C'_0 + C''_0}}, \quad (17)$$

where C' and C'' are the maximum and minimum values of backward-wave power at the detector in the directional coupler at resonance; C'_0 and C''_0 are the corresponding values in the absence of resonance.

Comparing the values of C' , C'' , C'_0 and C''_0 by means of the calibrated attenuator, on the basis of Equation (17) we can easily determine the reflection coefficient.

Since it is more difficult to satisfy conditions (16) than the analogous conditions (13), the accuracy in determining the reflection coefficient for the second variant will be lower than for the first variant.

4. SELECTION OF DIAPHRAGM TRANSFER CONSTANT AND LENGTH OF TEST PIECE

From Equations (7), (9) and (10) it is evident that the SWR (ρ and ρ_0) depend on the losses in the test piece and the end faces as well as on the transfer constant of the diaphragm. The minimum value of SWR must satisfy conditions (13) or (16) and, for convenience in measurement, the maximum value must not exceed 5-6. Let us discuss the sequence of operations in the selection of the diaphragm transfer constant and the length of the test piece, for example, for the case $\eta_0 < \eta_1 < \eta_B$. First the cavity is assembled without the test piece and, gradually increasing the diaphragm transfer constant, the SWR before the cavity is measured by the method described above. A typical dependence of SWR ρ_0 on the diaphragm transfer constant for the case $\eta_0 = 0.002$ is shown in Figure 4. From this curve it is seen that with an increase in the transfer constant the SWR at first decreases and, upon reaching unity, begins to increase.

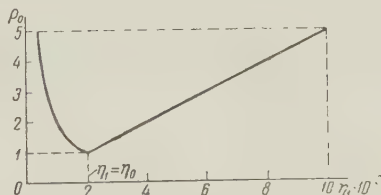


Figure 4. Standing-wave ratio as a function of diaphragm transfer constant.

The transfer constant η_1 must be so chosen that the value of ρ_0 is the maximum permissible value (the right hand branch of the curve in Figure 4).

The length of the test piece is determined by the condition that the SWR in front of the cavity with the test piece inserted is of the same order as without the test piece (in gradually increasing the length of the test piece, for the chosen transfer constant η_1 the SWR will decrease at first and, upon reaching unity, begin to increase).

It frequently occurs in practice that the length of the test piece is given and cannot be varied. In this case the diaphragm transfer constant and initial attenuation at the end faces η_0 are so chosen that the SWR, not exceeding 5-6, satisfies conditions (13) or (16).

5. THE DIAPHRAGM AND ITS CALIBRATION

The diaphragm is a thin waveguide partition with one (Figure 5a) or several (Figure 5b, c) apertures. The number of circular apertures or slits is so chosen that upon incidence of the H_{01} wave within the cavity no asymmetrical waves are excited. However, such diaphragms may, in addition to the H_{01} wave, give rise to symmetrical magnetic waves of higher order. Hence, diaphragms a and b may be used in the case where within the test waveguide there may be propagated a single symmetrical magnetic H_{01} wave ($Ka < 7.02$), while diaphragm c may even be used in the case where within the waveguide, in addition to the H_{01} wave, there may be propagated a symmetrical magnetic wave of the second order, that is, H_{02} ($Ka < 10.2$). The coupling apertures are located on the radius where currents induced by the H_{02} wave are equal to zero. Here K is the free-space propagation constant and a is the waveguide radius.

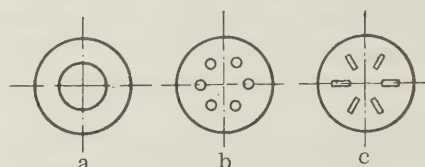


Figure 5. Various diaphragms.

Calibration of a diaphragm consists of determining the transfer constant η_1 within a range of frequencies. The block diagram of the setup for diaphragm calibration is given in Figure 6.

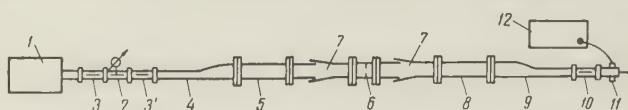


Figure 6. Block diagram of setup for diaphragm calibration.

The signal from klystron generator 1, modulated by rectangular pulses, is applied through decoupling attenuators 3 and 3' and calibrated attenuator 2 to an H_{01} wave exciter in circular waveguide 4. The H_{01} wave, passing through helical filter 5, reaches diaphragm 6, on both sides of which are located telescoping junctions 7 which permit eliminating reflection from the generator and receiving ends of the setup. Then (through helical filter 8, second exciter 9 and decoupling attenuator 10) the signal reaches detector 11 and, finally, measuring amplifier 12.

The transfer constant of the diaphragm is determined by comparing the signal level at detector 11 in the presence of the diaphragm with the signal level at detector 11 with the diaphragm removed. It is easily seen that reflections from both exciters introduce errors in determining the transfer constant.

In order to eliminate this error telescoping junctions 7 are used. With changes in distance from the diaphragm location to the H_{01} wave exciter the amplitude of the wave reaching detector 11 varies from

$$u_{\max} = \frac{A}{1 - k_1 k_2} \quad \text{to} \quad u_{\min} = \frac{A}{1 + k_1 k_2} \quad (18')$$

in the absence of the diaphragm and from

$$u_{d \max} = \frac{A \sqrt{\eta_2}}{1 - k_1 - k_2 + k_1 k_2} \quad \text{to} \quad u_{d \min} = \frac{A \sqrt{\eta_1}}{1 + k_1 + k_2 + k_1 k_2} \quad (18'')$$

in the presence of the diaphragm, where A is a constant, $k_1 \ll 1$ and $k_2 \ll 1$ are the absolute values of the coefficient of reflection from the exciters*.

* k_1 and k_2 represent the absolute values of the total reflection coefficients from the direction of the circular waveguide as caused by the presence of the H_{01} exciter and incomplete matching of the decoupling attenuators 3' and 10 with the line.

On the basis of Equations (18), ignoring second order terms, we can write the following expression for the transfer constant of the diaphragm:

$$\eta = \frac{1}{4} \left(\frac{u_{d \max}}{u_{\max \min}} + \frac{u_{d \min}}{u_{\max \min}} \right)^2. \quad (19)$$

Thus, in order to determine the transfer constant it is sufficient to compare the amplitudes of $u_{d \max}$ and $u_{d \min}$ with $u_{\max} \approx u_{\min}$; this comparison is performed by means of the calibrated attenuator.

CONCLUSION

The proposed method for the measurement of attenuation permits considerable simplification of the measurement setup. In using this method it is not necessary to measure the width of the resonance curve, which eliminates the need for high signal generator stability, high linearity of the oscillograph sweep, etc. Although the setups described above dealt with the testing of circular waveguide elements, the proposed method can be used in measuring losses in waveguides of other configurations.

REFERENCES

1. M.S. Khaykin, Measurement of surface resistance of superconducting tin at 9380 Mc, Dokl. AN SSSR, 1950, 75, 5, 661.
2. V.M. Vakhnin, T.F. Kolodina, Setup for measuring H_{01} wave attenuation in short waveguide sections by the resonant cavity method. Radiotekhnika i elektronika, 1956, 1, 12, 1485.
3. Microwave Measurements, Izd. Sovetskoye radio, 1952.
4. Yu.N. Kazantsev, Cavity resonator calculations by the method of superposition of electromagnetic waves, Radiotekhnika i elektronika, 1959, 4, 9, 1480.
5. J.K. Southworth, Principles and Application of Waveguide Transmission, Izd. Sovetskoye radio, 1955.
6. Yu.M. Isayenko, Tapered H_{01} wave exciter in circular waveguide, Radiotekhnika i elektronika, 1959, 4, 8, 1398.
7. M.V. Persikov, Directional coupler for H_{01} waves in circular waveguide, Radiotekhnika i elektronika, 1957, 2, 1, 65.
8. S.E. Miller, Coupled wave theory and waveguide applications, Bell System Techn. J., 1954, 33, 3, 661.
9. Yu.N. Kazantsev, V.V. Meriakri, Attenuation measurement in circular waveguides, Radiotekhnika i elektronika, 1959, 4, 1, 131.
10. Yu.N. Kazantsev, Yu.I. Kaznacheyev, V.V. Meriakri, Investigation of helical waveguides, Radiotekhnika i elektronika, 1959, 4, 11, 1816.
11. Yu.N. Kazantsev, V.V. Meriakri, Transmission of H_{01} waves through smooth bends, Radiotekhnika i elektronika, 1959, 4, 1, 133.
12. J. Storer, L. Sheyngold, S. Stein, Simple graphical analysis of a waveguide junction with two outputs, Sb. Vopr. radiolokatsionnoy tekhniki (Radar engineering symposium), 1954, 4, 109.

Institute of Radio Engineering
and Electronics, AN SSSR

Submitted to the editors 7 July 1959
After revision 11 February 1960

SOME PROBLEMS OF MICROWAVE CHANNEL MATCHING IN RADIO SPECTROSCOPES

V.I. Muromtsev and A.K. Piskunov

The paper discusses special aspects of matching the microwave channel of a radio spectro-scope with a reflex resonator. Analyses are given for possible nonlinear distortions of the paramagnetic absorption signal which arise under certain conditions in the matching of the resonator with the waveguide system. The influence of the dispersion signal on the shape of the absorption signal is evaluated.

INTRODUCTION

Analysis of the operation of various types of radio spectroscopy has been presented in References 1-3. In Reference 1 the chief causes of limited sensitivity of spectroscopes were examined in detail and the optimum conditions for matching of the microwave channel with the resonator were determined. The theory of simple and superheterodyne spectroscopes and the influence of dispersion on the observed shape of the paramagnetic absorption signal in a re-entrant resonator were first discussed in Reference 3.

The present paper presents a detailed analysis of some specific aspects of matching of the microwave system of spectroscopes with the reflex resonator, possible nonlinear distortions of the signal shape, and evaluation of the influence of the dispersion signal on the shape of the absorption signal.

In order to increase spectroscopy sensitivity it is necessary to match the microwave wave-guide system with the resonator. The optimum matching condition depends not only on the type of spectroscopy and the detector characteristics, but also on the internal noise of the generator, the noise figure of the amplifier and the noise temperature of the detector or converter.

It is most important to know the optimum matching conditions from the point of view of obtaining the maximum signal-to-noise ratio. As a first rough approximation it is often useful to know the optimum matching condition at which the maximum signal is obtained at the output of the system. Calculation of the matching conditions is based on the assumption that the waveguide system with a resonator is described by an equivalent lossless long line with characteristic resistance R_0 loaded by a lumped impedance Z_L .

The resonator is represented as an equivalent tuned circuit and is characterized by an equivalent loss resistance r , Q_0 and resonant frequency f_0 . The coupling of the resonator with the system is represented by an ideal transformer with transformation ratio n .

The lumped impedance loading the long line is defined as the input resistance of the equivalent circuit shown in Figure 1.

The reflection coefficient Γ is defined by

$$\Gamma = \frac{U_{\text{ref}}}{U_{\text{inc}}} = \frac{Z_L - R_0}{Z_L + R_0}, \quad (1)$$

where U_{ref} is the amplitude of the reflected wave and U_{inc} is the amplitude of the incident wave.

During observation of the absorption signal the resonator is tuned to the generator frequency. In this case the load impedance Z_L of the long line will be resistive and equal to

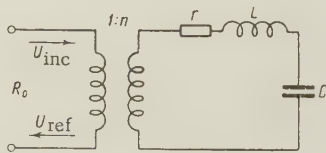


Figure 1. Equivalent circuit of waveguide system with resonator.

$$Z_L = R_L = \frac{r}{n^2}. \quad (2)$$

With tuning of the resonator to the generator frequency the reflection coefficient Γ is defined by

$$\Gamma_p = \frac{r - R_0 n^2}{r + R_0 n^2}. \quad (3)$$

Paramagnetic absorption of the microwave power causes additional losses in the resonator, which is the same as an increase in the equivalent resistance r , characterizing the loss, by Δr .

As the result of the change in loss there is a change in amplitude of the reflected wave. From Equations (1) and (3) it follows that

$$U_{\text{ref}} = \frac{r - R_0 n^2}{r + R_0 n^2} U_{\text{inc}} \quad (4)$$

or

$$U_{\text{ref}} = f(r, n). \quad (5)$$

Assuming that the paramagnetic losses are small, the magnitude of change in amplitude of the reflected wave may be written in the form

$$\Delta U_{\text{ref}} = \frac{\partial f}{\partial r} \Delta r$$

or

$$\Delta U_{\text{ref}} = 2 \frac{R_0 n^2}{(R_0 n^2 + r)^2} \Delta r U_{\text{inc}}. \quad (6)$$

As follows from Equation (6), at one and the same value of paramagnetic absorption (Δr) the change in amplitude of the reflected wave depends on the transformation ratio n^2 .

The magnitude of the signal ΔU_{ref} will be a maximum when

$$\frac{R_0 n^2}{r} = 1 \quad (7)$$

and is given by

$$\Delta U_{\text{ref}} = \frac{1}{2} U_{\text{inc}} \frac{\Delta r}{r}. \quad (8)$$

In practice it is more convenient to express ΔU_{ref} and the optimum coupling condition (7) as functions of Γ_p . Replacing $R_0 n^2$ in Equation (6) by the reflection coefficient Γ_p , we obtain from Equation (3)

$$\Delta U_{\text{ref}} = \frac{1}{2} U_{\text{inc}} (1 - \Gamma_p^2) \frac{\Delta r}{r}. \quad (9)$$

Equation (7), defining the condition at which the change in amplitude of the reflected wave will be maximum is written in the form

$$\Gamma_p = 0. \quad (10)$$

Condition (10) defines the optimum coupling in observing the absorption signal in a reflection spectroscope with linear detection.

1. MATCHING IN SQUARE-LAW DETECTION

It is known that the matching condition at which the maximum signal is obtained at the output depends on the detector characteristic. In the general case the choice of the matching condition depends also on the level of the incident power at the resonator.

Let us determine the condition of optimum matching in square-law detection in a circuit of the reflex type in observing an absorption signal.

In the spectroscopic case the change in the amplitude of the microwave carrier represents the electron spin resonance signal. The voltage applied to the detector is

$$U_{\text{inc}} = (U_{\text{ref}} + \Delta U_c) \cos \omega t, \quad (11)$$

where U_{ref} is the carrier amplitude and ΔU_c is the change in amplitude of the microwave voltage caused by paramagnetic absorption. The expression for the signal component after the square-law detector has the form

$$I_c = \alpha [(U_{\text{ref}} + \Delta U_c) \cos \omega t]^2 = \alpha U_{\text{ref}} \Delta U_c. \quad (12)$$

Thus, the value of the signal at the output is proportional to the amplitude of the carrier U_{ref} and the value of the signal ΔU_c at the input. Here, for an increase in the transfer constant of the detector at the same signal ΔU_c it is necessary to increase the carrier amplitude. Spectroscopes in which in order to increase the detector transfer constant, the carrier level is artificially increased are described, for example, in Reference 2. For a spectroscope with a reflex resonator, from Equations (1) and (9), we have

$$U_{\text{ref}} = \Gamma_p U_{\text{inc}},$$

$$\Delta U_c = \Delta U_{\text{ref}} = (1 - \Gamma_p^2) U_{\text{inc}} \frac{\Delta r}{2r}. \quad (13)$$

Substituting Equation (13) in Equation (12), we obtain the dependence of the signal at the output on the reflection coefficient Γ_p :

$$I_c = \text{const } \Gamma_p (1 - \Gamma_p^2) U_{\text{inc}}^2. \quad (14)$$

Analysis of Equation (14) shows that with

$$\Gamma_p = \pm \frac{1}{\sqrt{3}} \quad (15)$$

the signal at the output will be maximum. A plus sign before Γ_p indicates coupling below critical the negative sign indicates coupling above critical. In square-law conversion the signal at the output is proportional to the square of the voltage U_{inc}^2 , that is, the incident power at the resonator. In spectroscopes employing crystal detectors the optimum matching condition (15) is applicable at powers not greater than a fraction of a milliwatt. At high powers Equations (14) and (15) are not applicable.

We will note that with $|\Gamma_p|$ less than optimum the decrease in signal at the output is caused by a decrease in the transfer constant of the detector, and at values of $|\Gamma_p|$ greater than optimum; a decrease in signal at the output is associated with a decrease in the absorption signal ΔU_c , which is caused by a decrease in the microwave power entering the resonator. Matching ensuring maximum spectroscopic sensitivity differs from condition (15), which was pointed out in Reference 3.

Results of the investigation performed may be briefly formulated as follows:

1. The optimum matching condition in square-law detection is defined by Equation (15).
2. For optimum square-law conversion the optimum matching conditions do not depend on the level of power incident at the resonator.
3. The signal at the output of the square-law detector is proportional to the incident power at the resonator.

2. MATCHING IN A SIMPLE REFLECTION SPECTROSCOPE AT MEDIUM AND HIGH POWER LEVELS

In the operation of a simple spectroscope with nonsaturable specimens, in order to increase sensitivity it is necessary to increase the power level of the generator. At powers of the order of several milliwatts, depending on Γ_p , the detector may operate both in the square-law region and in the linear region. Let us discuss the manner in which in this case it is necessary to select the reflection coefficient Γ_p . Over a wide range of input voltages the detector characteristic has a relatively complex shape. Analytical determination of the optimum matching conditions is an extremely laborious undertaking. Hence we shall limit the discussion to a qualitative examination of the matching conditions at which maximum spectroscope sensitivity is obtained.

In the case under discussion the selection of the reflection coefficient Γ_p is based on the following considerations. In the square-law region of the detector characteristic the transfer constant is proportional to the carrier amplitude. With transition to the intermediate region the increase in transfer constant is retarded and with further increase in amplitude (into the

linear region) the transfer constant is practically independent of carrier amplitude.

However, in linear operation with an increase in voltage applied to the detector there is a sharp increase in anomalous noise of the detector, particularly in the audio-frequency region. At high levels of incident power at the detector even a small deviation of Γ_p from zero may lead to transition from square-law detection to linear detection. Consequently, in contrast to the preceding case, the optimum matching condition is determined, on the one hand, by a decrease in transfer constant upon a decrease in carrier level and, on the other hand, by an increase in anomalous noise of the detector with an increase in the applied voltage. The choice of Γ_p is determined by proper choice of detection conditions. Optimum detection conditions depend on the noise figure of the amplifier and on the individual properties of the crystal detector. For a simple spectroscope the frequency and amplitude noise of the generator may be disregarded.

Experiments reveal that for spectroscopes with modulation of the magnetic field no higher than 1 kc the optimum operating conditions lie between the square-law and linear regions.

Thus, in a simple spectroscope operating at various power levels, in order to ensure optimum detection conditions it is necessary to select appropriate values of reflection coefficient Γ_p .

3. DISTORTION OF ABSORPTION SIGNAL SHAPE IN THE MICROWAVE SYSTEM

Experience in operation with a simple reflex circuit (particularly with the superheterodyne circuit) shows that with the reflection coefficient Γ close to zero, the shape of the absorption signal is distorted due to the presence of the dispersion signal, even with small deviation of the generator frequency from the natural frequency of the resonator. A detailed discussion of this problem is presented below. For large absorption signals distortion of shape is observed even with precise tuning of the generator to the resonant frequency of the resonator. These distortions are not associated with the influence of dispersion.

Experimentally observed distortions of the electron spin resonance signal shape of diphenylpicryl-hydrazyl depending upon matching conditions with precise tuning of the generator frequency are shown in Figures 2-4. The signals were observed on a superheterodyne radio spectroscope similar to that described in Reference 3.

As is seen from these figures, at large signal levels the shape of the absorption curve may be considerably distorted. Let us discuss the cause of the observed distortion. The matching condition $\Gamma_p = 0$ corresponds to the case of critical coupling. Departure of the reflection coefficient from zero may occur in the above-critical direction and in the below-critical direction. With the reflection coefficient close to zero the amplitude of the reflected wave is defined by the approximate expression

$$U_{\text{ref}} \simeq (r - R_0 n^2) \frac{U_{\text{inc}}}{2r}. \quad (16)$$

With matching conditions below critical ($R_0 n^2 / r < 1$) and at critical ($R_0 n^2 / r = 1$) Equation (16) is conveniently written in the form

$$U_{\text{ref}} = \Delta R_L \frac{U_{\text{inc}}}{2r}, \quad (17)$$

where

$$\Delta R_L = r - R_0 n^2 \gg 0.$$

Paramagnetic absorption causes an increase in resistance ΔR_L by Δr and, consequently, a corresponding increase in the amplitude of the reflected wave

$$U_{\text{ref}} = (\Delta R_L + \Delta r) \frac{U_{\text{inc}}}{2r}. \quad (18)$$

The increase in the amplitude of the reflected wave is the signal. The magnitude of the signal is proportional to Δr and, consequently, with matching conditions below critical and at critical the shape of the signal is not distorted.

With matching conditions above critical the paramagnetic absorption will cause a change in amplitude of the reflected wave which is defined by the expression

$$U_{\text{ref}} = |\Delta R_B - \Delta r| \frac{U_{\text{inc}}}{2r}, \quad (19)$$

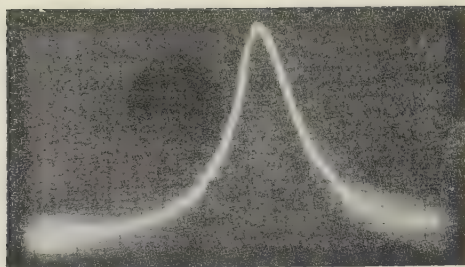


Figure 2. Absorption signal of diphenylpicryl-hydrazyl at below-critical coupling.

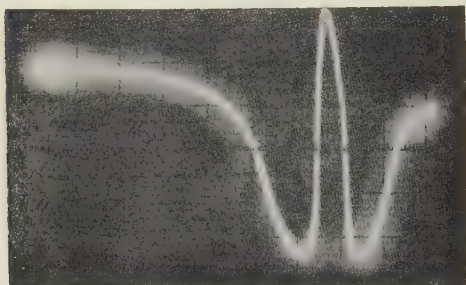


Figure 3. Absorption signal of diphenylpicryl-hydrazyl at above-critical coupling with $\Delta R_{B1} < \Delta r_{\max}$.

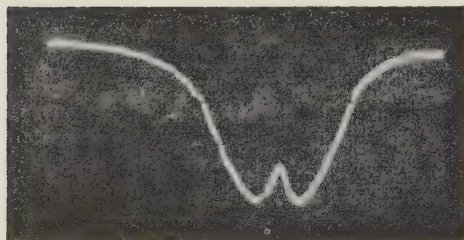


Figure 4. Absorption signal of diphenylpicryl-hydrazyl at above-critical coupling with $\Delta R_{B2} > \Delta R_{B1}$ but $\Delta R_{B2} < \Delta r_{\max}$.

where

$$\Delta R_B = |r - R_0 n^2|.$$

As follows from Equation (19), with an increase Δr the amplitude of U_{ref} at first decreases, at $\Delta R_B = \Delta r$ becomes equal to zero, and with a further increase in Δr begins to increase.

Thus, in the case of above-critical coupling the proportionality between the paramagnetic absorption of microwave power and the change in amplitude of the reflected wave may be disturbed and the shape of the signal distorted. In order that the signal shape may not be

distorted at above-critical coupling it is necessary to increase the mismatch (ΔR_B) or to decrease the size of the specimen.

Let it be noted that, in accordance with Equations (18) and (19), in the transition from below-critical matching to above-critical matching the signal phase is reversed. Equations (18) and (19) clearly define the observed signals. The signal in Figure 2 corresponds to the case of below-critical matching. The signals in Figures 3 and 4 correspond to the case of above-critical matching, wherein $|\Gamma_p|$ for the signal in Figure 3. In both cases (Figures 3 and 4) $\Delta R_B < \Delta r_{\max}$.

4. AMPLITUDE OF REFLECTED WAVE AS A FUNCTION OF DETUNING $\Delta\omega$ UNDER VARIOUS MATCHING CONDITIONS

The amplitude of the reflected wave is defined by the expression

$$U_{\text{ref}} = \left| \frac{Z_L - R_0 n^2}{Z_L + R_0 n^2} \right| U_{\text{inc}}, \quad (20)$$

where

$$Z_L = r + i \left(\omega L - \frac{1}{\omega C} \right).$$

Letting $X = \omega L - \frac{1}{\omega C}$, the modulus of the reflection coefficient may be written in the form

$$|\Gamma| = \sqrt{\frac{(r - R_0 n^2)^2 + X^2}{(r + R_0 n^2)^2 + X^2}}. \quad (21)$$

As is known [4], for frequencies near resonance

$$\frac{X}{r} = Q_0 \frac{2\Delta\omega}{\omega_0} = \frac{\Delta\omega}{\Delta\omega_0}, \quad (22)$$

where ω_0 is the resonant frequency, Q_0 is the Q of the unloaded resonator, $\Delta\omega_0$ is one-half the total 3-db bandwidth; $\Delta\omega = \omega - \omega_0$ is the detuning from resonance. Applying Equation (22) the modulus of the reflection coefficient is conveniently expressed in the form

$$|\Gamma| = \sqrt{\frac{\left(1 - \frac{R_0 n^2}{r}\right)^2 + \left(\frac{\Delta\omega}{\Delta\omega_0}\right)^2}{\left(1 + \frac{R_0 n^2}{r}\right)^2 + \left(\frac{\Delta\omega}{\Delta\omega_0}\right)^2}}. \quad (23)$$

It follows from Equation (3) that

$$\frac{R_0 n^2}{r} = \frac{1 - \Gamma_p}{1 + \Gamma_p}. \quad (24)$$

Substituting Equation (24) in Equation (23), we obtain the final expression for dependence of the modulus of the reflection coefficient $|\Gamma|$ on detuning under various coupling conditions:

$$|\Gamma| = \sqrt{\frac{\left(\frac{2\Gamma_p}{1 + \Gamma_p}\right)^2 + \left(\frac{\Delta\omega}{\Delta\omega_0}\right)^2}{\left(\frac{2}{1 + \Gamma_p}\right)^2 + \left(\frac{\Delta\omega}{\Delta\omega_0}\right)^2}}. \quad (25)$$

Equation (25) is of primary importance in evaluating the influence of the dispersion signal and the frequency stability of the generator under various matching conditions in the reflex circuit of the spectroscopy.

Curves for the dependence of $|\Gamma|$ on the relative detuning $y = \Delta\omega/\Delta\omega_0$ at values of $\Gamma_p = 0, 0.1$ and 0.3 are given in Figure 5.

As is seen from Figure 5, at a given value of detuning y the change in value of $|\Gamma|$ and, consequently, in the amplitude of the reflected wave will be maximum with $|\Gamma_p| = 0$. This means that the value of noise caused by frequency instability of the generator at $\Gamma_p = 0$ will be maximum. If the dispersion of frequency instability is small in comparison with the bandwidth of the resonator

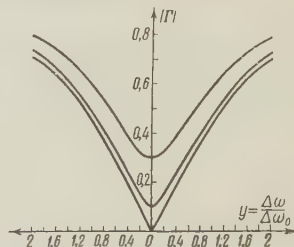


Figure 5. Dependence of the modulus of the reflection coefficient on detuning at various values of Γ_p .

then an increase in $|\Gamma_p|$ to 0.1 - 0.2 substantially decreases the frequency noise.

5. MATCHING WITH $|\Gamma|$ CLOSE TO ZERO

In the observation of the absorption signal in a superheterodyne circuit the value of the reflection coefficient Γ_p is chosen close to zero.

Let us examine the modulus of the reflection coefficient as a function of detuning y at values of Γ_p lying within the range from zero to 0.3.

We shall limit the discussion to cases of small detuning:

$$y = \left| \frac{\Delta\omega}{\Delta\omega_0} \right| \leq 0.2. \quad (26)$$

In this case

$$\left(\frac{2}{1 + \Gamma_p} \right)^2 \gg \left(\frac{\Delta\omega}{\Delta\omega_0} \right)^2 \quad (27)$$

and Equation (25) may be written in the form

$$|\Gamma| = b \sqrt{a^2 + y^2}, \quad (28)$$

where

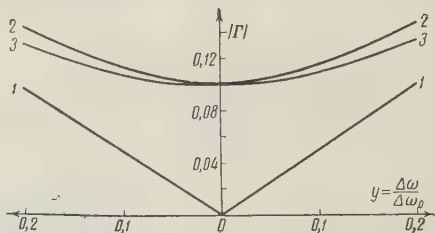
$$b = \frac{1 + \Gamma_p}{2}; \quad a = \frac{2\Gamma_p}{1 + \Gamma_p}.$$

Curves for the dependence of $|\Gamma|$ on detuning under various matching conditions are given in Figure 6.

It is seen from the curves that at the same value of $|\Gamma|$ and the same relative detuning y the value of the modulus of the reflection coefficient at above-critical coupling is smaller than at below-critical coupling.

Figure 6. Behavior of the modulus of the reflection coefficient at small values of detuning y and at above-critical and below-critical coupling.

1 - Critical matching ($\Gamma_p=0$); 2 - below-critical matching ($\Gamma_p=0.1$); 3 - above-critical matching ($\Gamma_p=0.1$).



Let Γ_B represent the reflection coefficient at below-critical coupling, Γ_A the reflection coefficient at above-critical coupling, and

$$a_B = \frac{2|\Gamma_p|}{1 + |\Gamma_p|}, \quad a_A = \frac{2|\Gamma_p|}{1 - |\Gamma_p|}. \quad (29)$$

The ratio of the modulus of the reflection coefficient at below-critical coupling to the modulus of the reflection coefficient at above-critical coupling may be written in the form

$$\frac{|\Gamma_B|}{|\Gamma_A|} = \sqrt{\frac{1 + \left(\frac{y}{a_B}\right)^2}{1 + \left(\frac{y}{a_A}\right)^2}}. \quad (30)$$

Since, in accordance with Equations (29), $a_A > a_B$, then from Equation (30) it follows that

$$|\Gamma_B| > |\Gamma_A|. \quad (31)$$

It can be shown that for $y \neq 0$

$$\frac{d|\Gamma_B|}{dy} \bigg|_{\frac{d|\Gamma_A|}{dy}} = \frac{\sqrt{a_A^2 + y^2}}{\sqrt{a_B^2 + y^2}}, \quad (32)$$

from which it is seen that

$$\frac{d|\Gamma_B|}{dy} > \frac{d|\Gamma_A|}{dy}. \quad (33)$$

Inequality (33) indicates that at above-critical coupling frequency instability of the generator has less effect than at below-critical coupling.

Substituting into Equation (32) the values of a_A and a_B at various Γ_p and various detuning y , we may obtain a quantitative determination of the decrease in slope and, consequently, the decrease in the influence of frequency noise in the transition from below-critical matching to above-critical matching. The expression for the slope of the reflection curve $d|\Gamma|/dy$ at small detuning values, utilizing Equation (28), has the form

$$\frac{d|\Gamma|}{dy} = \pm b \frac{y}{\sqrt{a^2 + b^2}}. \quad (34)$$

It follows from Equation (34) that with a decrease in $|\Gamma|$ to zero the value of $d|\Gamma|/dy$ approaches $\pm 1/2$. With $|\Gamma_p|$ equal to zero, the value of $d|\Gamma|/dy = \pm 1/2$ for all values of y satisfying Equation (22). With $\Gamma_p = 0.1$ and $y = 0.05$ the value of $d|\Gamma|/dy = 0.15$. With $\Gamma_p = 0.3$ and $y = 0.05$ the value of $d|\Gamma|/dy = 0.075$. These examples show that at small detuning ($y = 0.05$) the slope of the reflection curve depends substantially on $|\Gamma_p|$. The curve for the dependence of the slope of the reflection curve $d|\Gamma|/dy$ on the value of $|\Gamma_p|$ at a relative detuning of $y = 0.02$ for below-critical coupling is shown in Figure 7. With $|\Gamma_p| < 0.2$ for above-critical coupling the derivative of the reflection curve is practically the same as the curve in Figure 7.

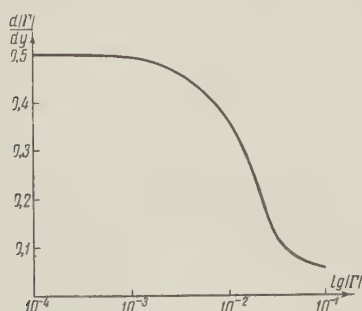


Figure 7. Dependence of slope of reflection curve $d|\Gamma|/dy$ on the value of $\lg |\Gamma_p|$ with relative detuning $y = 0.02$.

6. INFLUENCE OF THE DISPERSION SIGNAL ON THE SHAPE OF THE ABSORPTION SIGNAL

In observing the absorption signal it is desirable that the change in amplitude of the reflected wave, caused by paramagnetic absorption, be maximum and that the influence of paramagnetic dispersion be minimum. It was previously shown that if the chosen value of detuning $y = 0$, then the value of the absorption signal ΔU_c will be maximum under the condition $\Gamma_p = 0$. The equations

$$\Gamma_p = 0, \quad y = 0, \quad (35)$$

defining the optimum coupling and tuning in observing the absorption signal, were obtained without consideration of the influence of paramagnetic dispersion. The influence of dispersion is more evident at detuning $y \neq 0$. It is of interest to examine the influence of dispersion on the shape of the absorption signal at detuning $y = 0$.

Let us discuss this problem in the case of a reflex spectroscopy without automatic frequency control. Let us assume that in the absence of paramagnetic resonance $y = 0$. Paramagnetic dispersion causes a change in the energy of the magnetic field within the resonator, which is equivalent to a change in the resonator inductance L by the value ΔL .

Assume that paramagnetic absorption causes an increase in equivalent loss resistance

by a value of ΔL . This results in changes in Γ_p and y :

$$\Delta \Gamma_p = \frac{2R_0 n^2}{(R_0 n^2 + r)^2} \Delta r,$$

$$\Delta y = Q_0 \frac{\Delta L}{L} = \frac{\omega_0}{r} \Delta L. \quad (36)$$

The result is that the value of the reflection coefficient Γ is changed. Taking into account Equations (25), (35), (36) we obtain for $\Delta \Gamma$ the following approximate expression:

$$\Delta \Gamma \simeq \frac{1}{2} \sqrt{(2\Delta \Gamma_p)^2 + \Delta y^2} = \frac{1}{2r} \sqrt{\Delta r^2 + (\omega_0 \Delta L)^2}. \quad (37)$$

It follows from Equation (37) that the shape of the reflected signal under condition (35) is a complex function of the absorption and dispersion signals.

In a spectroscope with a reflex resonator and without automatic frequency control of the operating cavity we shall compare the contribution of dispersion to the observed paramagnetic resonance signal under condition (35) with the contribution of the absorption signal. We shall show that with an increase in $|\Gamma_p|$ at tuning of $y = 0$ the influence of the dispersion signal is substantially decreased. Let us examine the matching at which

$$|\Gamma_p| \gg |\Gamma_{\min}|, \quad y = 0, \quad (38)$$

wherein

$$|\Gamma_{\min}| \gg |\Delta y|, \quad |\Gamma_{\min}| \gg |\Delta \Gamma_p|. \quad (39)$$

Since the magnitudes of $|\Delta y|$ and $|\Delta \Gamma_p|$ are usually quite small, the value of $|\Gamma_{\min}|$ may be considerably less than 0.01. In the case defined by Equation (38).

$$\Delta \Gamma \simeq \frac{1}{2} \sqrt{4(\Gamma_{\min} + \Delta \Gamma_p)^2 + \Delta y^2} - |\Gamma_{\min}|. \quad (40)$$

Disregarding quantities of second order in Equation (40), we find that

$$\Delta \Gamma \simeq \Delta \Gamma_p.$$

Thus, with sufficiently large $|\Gamma_p|$ the influence of the dispersion signal on the shape of the absorption signal may be disregarded.

7. MATCHING IN SQUARE-LAW DETECTION IN THE PRESENCE OF FREQUENCY INSTABILITY OF THE GENERATOR

It has been shown that in square-law detection the signal magnitude after the detector will be a maximum when the reflection coefficient $\Gamma_p = \pm 1/\sqrt{3}$. If it is not possible to disregard the influence of frequency noise within the spectroscope, then the two cases of matching are not equivalent.

It can be shown that the slope of the reflection curve with $\Gamma_p = 1/\sqrt{3}$ and $y \leq 0.1$ is defined by the approximate expression

$$\frac{\partial |\Gamma_p|}{\partial y} \simeq 0,7 y.$$

With $\Gamma_p = -1/\sqrt{3}$ the slope of the reflection curve has the form

$$\frac{\partial |\Gamma_A|}{\partial y} \simeq 0,05 y.$$

The slope of the reflection curve at below-critical matching is approximately 14 times greater than at above-critical matching. Usually the ratio of the dispersion of frequency instability to the one-half bandwidth of the resonator is much less than unity. Hence we may approximately consider that the dispersion of amplitude instability at the detector input is proportional to the slope of the reflection curve. This indicates that in the case of square-law detection at above-critical matching the frequency noise will be 14 times less than at below-critical matching.

CONCLUSION

In a spectroscope with a reflex resonator a decrease in $|\Gamma_p|$ leads to an increase in the slope of the reflection curve at small values of detuning y .

The increase in slope results in an increase in the effect of the dispersion signal on the shape of the absorption signal and there is also an increase in noise caused by frequency instability of the microwave oscillator.

With $|\Gamma_p| = 0$ and precise tuning at $y = 0$ the shape of the electron spin resonance signal is a complex function of the absorption and dispersion signal. In this case the noise caused by frequency instability is maximum.

The increase in frequency noise and the influence of the dispersion signal with a decrease in $|\Gamma_p|$ explains why superheterodyne radio spectroscopes do not need i-f amplifiers with extremely high gain, for an increase in i-f gain in ordinary superheterodyne spectroscopes implies a corresponding decrease in $|\Gamma_p|$.

With an increase in incident power at the resonator, in order to maintain the required operating conditions of the frequency converter in an ordinary superheterodyne spectroscope it is necessary to decrease $|\Gamma_p|$. This leads to an additional increase in the influence of frequency instability.

We wish to thank A. A. Manenkov for his valuable advice.

REFERENCES

1. G. Feher, Bell System Techn. J., 1957, 36, 2, 449.
2. D. I. E. Ingram, Free Radicals as Studied by Elektron Spin Resonance, London, 1958.
3. A. A. Manenkov, A. M. Prokhorov, Radiotekhnika i elektronika, 1956, 1, 4, 469.
4. B. P. Aseyev, Oscillatory Circuits, Svyaz'izdat, 1955.

Submitted to the editors 19 May 1960.

LARGE-SIGNAL OPERATION OF BACKWARD-WAVE AMPLIFIER

G. N. Rapoport

On the basis of the numerical solution of nonlinear equations this paper investigates the amplification of signals of finite amplitude by a single-stage O-type BWT. It is shown that nonlinear effects in the BWT are much greater than in an ordinary TWT and practically exclude application of small-signal theory to the analysis of a BWT power amplifier. The paper also presents a brief discussion of the operating conditions in which self-oscillations are locked by a periodic signal at the BWT input.

SINGLE-STAGE AMPLIFIER

Operation of the O-type BWT as an amplifier was first reported by R. Kompfner [1]. The small-signal theory of BWT amplifiers has been elaborated in References 2 and 3. It has been established that up to the self-excitation threshold the BWT is a regenerative amplifier with relatively narrow bandwidth, which may be shifted to wider limits by varying the accelerating voltage. The small-signal power gain in the middle of the G band is determined from the approximate equation

$$G = \frac{1}{\left(1 - \frac{I_0}{I_K}\right)^2}, \quad (1)$$

where I_0 is the electron beam current and I_K is the critical current corresponding to the self-excitation threshold. In approaching the self-excitation threshold the gain rises, the bandwidth Δf decreases and the produce $\Delta f \sqrt{G}$ remains almost constant.

The present paper discusses the calculation of the gain and efficiency of a backward-wave amplifier taking into consideration nonlinear effects and establishes the limits of applicability of linear theory (small-signal theory).

As is known, for conventional TWT's substantial departures from linear theory occur only at output powers greater than 0.4 - 0.6 of saturation power. Hence small-signal theory is applicable in practical TWT calculations.

In conjunction with the regenerative nature of the amplification in a BWT it is to be expected that nonlinear effects will be of greater significance, particularly as beam current approaches the critical value.

The calculations presented below are based on an approximate model and the system of BWT equations discussed in References 4 and 5. For the sake of simplification the problem of the influence of the space charge was not considered. In this case the system of BWT equations has the form

$$\frac{d\bar{F}}{d\bar{x}} - \delta\bar{F} = -\frac{1}{\pi} \int_0^{2\pi} e^{iz(\alpha, \bar{x})} d\alpha, \quad (2)$$

$$\frac{\partial^2 z(\alpha, \bar{x})}{\partial \bar{x}^2} = -\left(1 - C \frac{\partial z(\alpha, \bar{x})}{\partial \bar{x}}\right)^3 \operatorname{Re} \{e^{iz(\alpha, \bar{x})} \bar{F}(\bar{x})\}, \quad (3)$$

where $\bar{F}(\bar{x})$ is the dimensionless amplitude of the high-frequency field, wherein

$$\bar{F}(\bar{x}) = \frac{v}{u} F(x); \quad \frac{|F|^2}{2} = \frac{P}{CI_0 V}; \quad (4)$$

P is the energy flux of the field; C is the gain parameter; I_0 is the beam current; V is the accelerating voltage; v is the initial electron velocity; $u = u(\omega)$ is the backward-wave phase velocity in a "cold" retarding system at frequency ω ;

$$\bar{x} = \frac{v}{u} x; \quad x = C \frac{\omega}{v} s;$$

s is the electron path; δ is the loss parameter; $Z(\alpha, \bar{x})$ is the electron phase relative to the wave in a cold system at point \bar{x} ; $\alpha = -Z(\alpha, 0)$ is the initial electron phase.

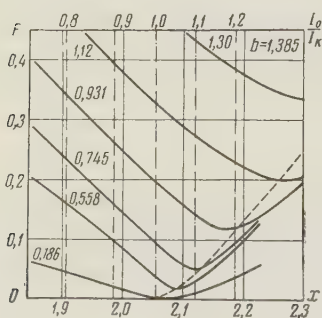


Figure 1. Distribution function of field amplitudes for various amplitudes at tube output. Parameter $b = 1.385$.

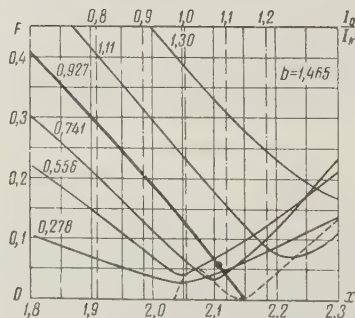


Figure 2. Distribution function of field amplitudes for various amplitudes at tube output. Parameter $b = 1.465$.

Equation (2) describes the excitation of the retarding system by the electron stream and Equation (3) describes the movement of electrons within the high-frequency field. If the wave dispersion $u(\omega)$ in the retarding system is known, then the frequency ω is uniquely associated with parameter

$$b = \frac{v - u}{Cv} = \left. \frac{\partial z}{\partial x} \right|_{x=0},$$

which, together with the amplitude of the field at the tube output $F(0)$, defines the initial conditions for the system of Equations (2)-(3).

With the frequency and amplitude of the wave at the output $F(0)$ so given, the amplitude of the wave at any point $s > 0$ is determined by the solution of the system of Equations (2)-(3).

For tubes with length of interaction space L the amplitude of the signal at the input will be $F(x_L)$, where $x_L = \frac{\omega}{V} L$.

The system of integro-differential equations in Equations (2)-(3) leads approximately to a system of 34 ordinary (nonlinear) differential equations of first order by replacing the continuous initial-phase distribution of electrons with a discrete distribution for 16 equally spaced initial phases. Integration was performed on the "Ural" electronic computer by the Runge-Kutta method of the fourth order.

The following values of the parameters were adopted in the calculations: $C = 0.05$, $\delta = 0.1$ ("cold" losses are approximately 0.27 db per length of retarded wave).

Figures 1 and 2 show the distribution functions of field amplitudes for various amplitudes of $F(0)$ at the output, the values of which are indicated on the curves. For Figure 1 the parameter $b = 1.385$, which corresponds to a signal frequency equal to the self-oscillation frequency of a BWT the length of which is brought to the threshold of self-excitation. For Figure 2, $b = 1.465$, which corresponds to a lower signal frequency equal to the natural frequency of self-oscillation of the BWT with a current ratio $I_0/I_k = 1.15$.

Values of x are plotted on the lower abscissa scale and the ratio of operating current to starting current (I_0/I_k) is plotted on the upper abscissa scale. Similar calculations were performed for a number of values of the parameter b from 1 to 2. In the region $x \leq 2.055$, where $I_0/I_k \leq 1$, the resulting solutions of $F(x)$ describe the BWT amplifier conditions. With $I_0/I_k > 1$, where $x > 2.055$, the resulting solutions of $F(x)$ correspond formally to conditions wherein self-oscillation is locked by an external signal.

At small output amplitudes (less than 0.3) the functions F approximate the functions calculated from small-signal theory. A characteristic nonlinear effect is the dependence of the position and value of the minimum of $F(x)$ on the amplitude of $F(0)$ at the output. The envelope of minima of $F(x)$ is represented by the broken line in Figures 1 and 2.

Let us discuss the principal characteristics of amplifier operation. Figures 3 and 4 show the dependence of gain G (in decibels) on the parameter b for various excitation powers P_{in} .

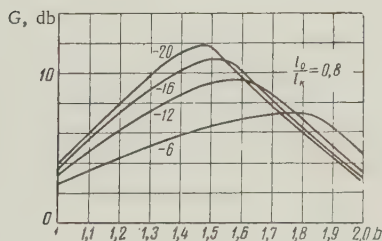


Figure 3. Gain G as a function of parameter b for various excitation powers P_{in} ; $I_0/I_k = 0.8$; numbers on curves are values of P_{in}/CI_0V (db).

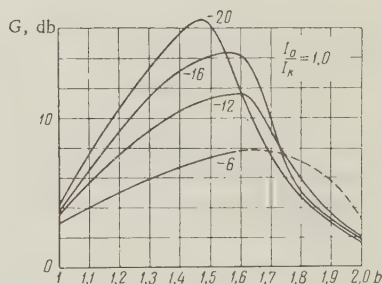


Figure 4. Gain G as a function of parameter b for various excitation powers P_{in} ; $I_0/I_k = 1.0$; numbers on curves are values of P_{in}/CI_0V (db).

From Figures 3 and 4 we can calculate the dependence of gain on the signal frequency f at a fixed accelerating voltage $V = \text{const}$ or the dependence of gain on V with $f = \text{const}$. Let

f_1 and f_2 represent frequencies corresponding to values of b_1 and b_2 with $V = \text{const}$. Then, as is easily seen,

$$\frac{f_2 - f_1}{\bar{f}} \approx \frac{-C}{f + \left| \frac{u_{ph}}{u_{gr}} \right|} (b_2 - b_1), \quad (5)$$

with $f = \text{const}$

$$\frac{V_2 - V_1}{\bar{V}} \approx 2C (b_2 - b_1), \quad (6)$$

where \bar{f} is the mean frequency, \bar{V} is the mean accelerating voltage, u_{ph} and u_{gr} are the phase and group velocities of the retarded (backward) wave.

It is seen from Figures 3 and 4 that with an increase in excitation power the maximum gain of the BWT decreases, the maximum of the gain curve is displaced in the direction of lower frequencies and the bandwidth increases. From Figures 3 and 4 and from Equation (5) it is seen that, with the given values of I_0/I_k and relative excitation power $P_{in}/C I_0 V$, in order to increase the bandwidth it is desirable to decrease the dispersion of the system and to increase the coupling of the system with the electron beam as well as the value of the beam current I_0 .

For signals with frequencies above or equal to the BWT self-oscillation frequency f_0 at the self-excitation threshold the gain decreases with an increase in signal power; however, at frequencies below f_0 a nonlinear increase in gain is possible with an increase in excitation power. An explanation of a similar effect in the TWT with large values of the parameter b is given in Reference 6.

Figure 5 shows the gain G (in decibels) as a function of the normalized excitation power (also expressed in decibels) for signal frequencies corresponding to maximum gain. For these same frequencies Figure 6 shows the output power $P_{out}/C I_0 V$ as a function of the excitation power. The broken curves in Figure 6 represent the difference $(P_{out} - P_{in})/C I_0 V$. For comparison in Figure 6 the arrow indicates the maximum BWT power with the given parameters C, δ, I_0, V in oscillator operation (the tube provides such power if the length of the retarding system is 40% greater than critical).

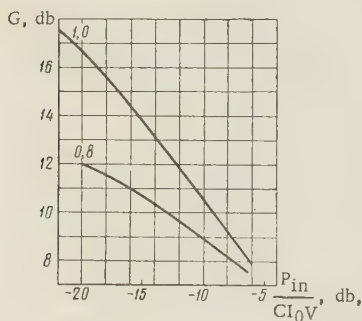


Figure 5. Gain G as a function of normalized excitation power for signal frequencies corresponding to maximum gain. $I_0/I_k = 0.8$ and $I_0/I_k = 1.0$.

From Figures 5 and 6 it is seen that with $I_0/I_k = 0.8$ the region of linear gain is limited to extremely small powers, not exceeding 10-15% of the saturation power (efficiency less than 1%). In approaching the self-excitation threshold the "linear" region is further diminished. It follows from this that small-signal theory is practically inapplicable for BWT power amplifier calculations. With $I_0/I_k = 1$ small-signal theory leads to infinite gain, whereas even at relatively low excitation power $P_{in} = 0.01 C I_0 V$ the actual gain is less than 17 db. M.S. Neyman has discussed the possibility of the synchronization of an O-type backward-wave tube by an external periodic signal. In Reference 7 he examines methods of synchronization and presents a qualitative characteristic of expectable dependences of the locking bandwidth on tube parameters and signal power.

Turning to an examination of locking conditions from the data of the calculations performed, we note that the subject of the stability of these conditions is not strictly within the province of this paper, for we proceed from assumptions of the stationary nature and single frequency nature of the oscillations. Nevertheless, with the information at hand we can easily confirm the existence of stable locking conditions.

The heavy line in Figure 2 represents the field distribution function $F_1(x)$ for the self-oscillating state of a tube with length $x_L = 2.15$ corresponding to $I_0/I_k = 1.15$. In this state the frequency of self-oscillation $f_0 = f_1$ corresponds to $b = 1.465$ and the amplitude at the output $F_1(0) = 0.927$. It is evident that a shorter tube, for example, $x_L = 2.12$ ($I_0/I_k = 1.08$)

for which the frequency of self-oscillation $f_0 = f_2 > f_1^*$, may be synchronized by a signal of frequency f_1 if the amplitude of the signal corresponds to the distribution function of the self-oscillating state and is $F_1(x_L) = 0.06$ (the point with coordinates x_L and $\bar{F} = F_1(x_L)$ in Figure 2, indicated by a dot). Since the self-oscillating state of a tube of length x_L is stable, the locking conditions of a tube of shorter length x_L^1 are also stable, which conditions are characterized by the same field distribution and current distribution as the self-oscillating state (heavy line in Figure 2).

The conditions of synchronization of a tube by signals of frequency f_1 but with large input amplitudes must also be stable. These conditions are represented by the descending branches of the distribution function which are located above the line of the self-oscillating condition F_1 and correspond to large output amplitudes.

From Figure 2 is seen that the lines of stable conditions intersect the ascending branches of the distribution function with output amplitudes $F(0) = 0.278-0.741$ smaller than for the self-oscillating state at the same frequency. Consequently, the ascending branches of the distribution function over the examined segment $2.055 < x_L < 2.15$ correspond to unstable conditions.

If in conjunction with the above discussion we assume that within the limits of stability of the locking conditions the BWT output amplitude must not be less than the amplitude of free self-oscillations F_a , then we may show that the change in parameter b within the locking bandwidth is $\approx 0.8 \frac{F_c}{F_a}$, where F_c is the amplitude of the synchronizing signal at the BWT input. Considering Equation (5), for a BWT with the above parameters we obtain the following evaluation of the width of the region Δf of frequency locking:

$$\frac{\Delta f}{f} \approx 0.8 \frac{\bar{C}}{1 + \frac{|u_{ph}|}{|u_{gr}|}} \frac{F_c^2}{F_a^2}. \quad (7)$$

It also follows from the above remarks that the self-oscillations of a BWT with finite output amplitude may be locked by as small a signal as is desired at the tube input if the frequency of the signal is close to the frequency of the self-oscillations.

The locking bandwidth of the tube, all other conditions being equal, will increase in inverse proportion to the dispersion of the retarding system. Let it be noted that the results obtained are in qualitative agreement with the conclusions presented in Reference 7.

A further problem in the theory is the examination of bi-frequency (asynchronous) conditions of a BWT and the determination of the limits of stability of synchronous operation. However, the results obtained also indicate, in principle, the possibility of use of a synchronized tube as a narrowband "amplifier" of modulated signals with a correspondingly limited spectrum.

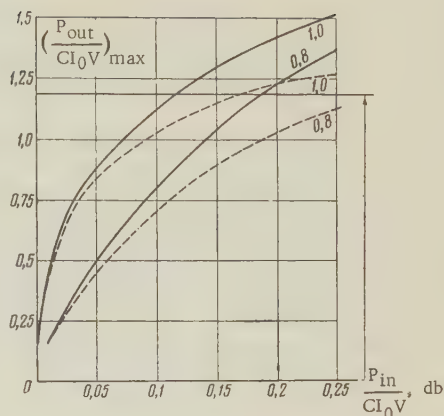


Figure 6. Output power P_{out}/CI_0V as a function of excitation power for signal frequencies corresponding to maximum gain. Broken curves in Figure 6 correspond to difference $(P_{out}-P_{in})/CI_0V$.

*With an increase in tube length or beam current the frequency of self-oscillation decreases [4].

REFERENCES

1. R. Kompfner, N.T. Williams, Backward-wave tubes, *Proc. IRE*, 1953, 41, 11, 1602.
2. H. Heffner, Analysis of the backward-wave travelling-wave tube, *Proc. IRE*, 1954, 42, 6, 930.
3. M.R. Currie, J.R. Whinnery, The cascade backward-wave amplifier: a high-gain voltage-tuned filter for microwaves, *Proc. IRE*, 1955, 43, 11, 1617.
4. G.N. Rapoport, Some results in the nonlinear theory of self-oscillation of a backward-wave tube with a longitudinal field, *Dokl. AN SSSR*, 1958, 117, 3, 411.
5. G.N. Rapoport, Nonlinear theory of the traveling-wave tube (summary), *Izv. vuzov MVO SSSR (Radiotekhnika)*, 1958, 5, 599.
6. L.A. Vaynshteyn, Nonlinear theory of the TWT. Part II, Numerical results, *Radiotekhnika i elektronika*, 1957, 2, 8, 1027.
7. M.S. Neyman, Some characteristics of the application of the method of forced frequency synchronization to self-excited backward-wave oscillators, *Izv. vuzov MVO SSSR (Radio-tekhnika)*, 1958, 3, 288.

Submitted to the editors 9 May 1960.

AMPLIFICATION OF SPACE-CHARGE WAVES IN THE PASSAGE OF ELECTRON BEAMS WITHIN MEDIA WITH INDUCTIVE SUSCEPTANCE

L. V. Kasatkin

On the basis of analysis of a one-dimensional system, this paper discusses the mechanism of interaction leading to the amplification of space-charge waves in the presence of inductive-resistive reaction of the medium beyond the natural passband. An approximate calculation of the nonlinearity of the bunching process is presented which permits evaluation of the gain characteristics.

The question of the amplification of space-charge waves in the passage of electron beams through a system with inductive susceptances or conductances relative to the beam has been discussed in a number of works.

In Reference 1 there is described a retarding system of the "double-comb" type and as the result of solution of the field equations it is shown that beyond the passband of this system, when the susceptance referred to the beam is inductive, the occurrence of a rising wave in the electron beam is possible. In Reference 2 the amplification characteristics of space-charge waves are determined (by a linear approximation) for different types of susceptance of the drift walls for the cases of cylindrical and plane symmetry. An extremely interesting application of this principle of amplification is described in Reference 3 where the so-called "inductive drift" (achieved in the form of a retarding system whose passband lies within a frequency region above that of the operating region) served as the active amplifying decoupler in a "hybrid" TWT. By using the amplification of the waves within the beam during its inductive drift passage it was possible to shorten the sections of the retarding waveguides of the buncher and of the collector and thereby to insure high stability of the amplifier at considerable gain and bandwidth.

The amplification mechanism of space-charge waves is not difficult to trace in the examination of a one-dimensional model in which the electron beam freely passes along thin channels in a medium having an inductive susceptance $y = -jB_L = -j/\omega L_1$ and also, in the general case, a conductance σ . The parameters of the medium are so chosen that in the

absence of an electron beam no electromagnetic waves may be propagated in the medium.

In the passage of an electron beam with variable components of current density i_d and charge density ρ_d through the system discussed there arises an induced charge with density ρ_c and current with density

$$i_c = \sigma E_z + \frac{1}{L_1} \int E_z dt. \quad (1)$$

Here the coefficient of induction L_1 characterizes the specific inductive reactance of the medium. During harmonic excitation the coefficient L_1 and the negative dielectric constant of the medium ϵ_c ($\epsilon_c < 0$) are related by the relationship $L_1 = -1/\omega^2 \epsilon_c$. The total charge density in the system

$$\rho_\Sigma = \rho_d + \rho_c, \quad (2)$$

and the total current density

$$i_\Sigma = \text{rot } H = i_d + \sigma E_z + \epsilon \frac{\partial E_z}{\partial t} + \frac{1}{L_1} \int E_z dt. \quad (3)$$

In determining the charge induced in the medium we shall use the equation

$$\text{div } E_z = \frac{\rho_\Sigma}{\epsilon} \quad (4)$$

and the discontinuity equation of beam convection current

$$\text{div } i_d + \frac{\partial \rho_d}{\partial t} = 0. \quad (5)$$

Applying the divergence operator to Equation (3), taking into account Equations (2), (4), (5) and the time function $e^{j\omega t}$, we obtain

$$\rho_c = -\rho_d \left[1 + \frac{1}{\frac{\sigma}{j\omega\epsilon} - \frac{B_L}{\omega\epsilon}} \right]^{-1}. \quad (6)$$

Depending on parameters σ , ϵ , L there arise in the medium induced charges, differing in phase and magnitude, the electric field of which interacts with the electron beam. The nature of this interaction may be determined by proceeding from the wave equation for the one dimensional system discussed:

$$\frac{\partial^2 E_z}{\partial z^2} - \mu\epsilon \frac{\partial^2 E_z}{\partial t^2} = \frac{1}{\epsilon} \nabla \rho_\Sigma + \mu \frac{\partial i}{\partial t}. \quad (7)$$

Here i is the current density, equal to the sum of the densities of the convection current within the beam and the current in the medium. Expressing ρ_c by ρ_d in accordance with Equation (6), expressing ρ_d and i by E_z (as was done in References 2 and 4) and assuming that all variable quantities depend on z and t according to the expression $e^{j(\omega t - \beta z)}$, we obtain instead of Equation (7)

$$E_z (\beta^2 - \omega^2 \mu\epsilon + j\omega\mu\sigma + \omega\mu B_L) \left[\frac{(\epsilon_0/\epsilon) \beta_p^2}{(\beta_e - \beta)^2 \left(1 + \frac{\sigma}{j\omega\epsilon} - \frac{B_L}{\omega\epsilon} \right)} - 1 \right] = 0, \quad (8)$$

where $\beta_p = \omega_p/u_0 = (\eta I_0/\epsilon u_0^3)^{1/2}$ is the wave number in the plasma, ϵ_0 is the dielectric constant in vacuum, I_0 is the constant component of beam current density, u_0 is the constant velocity of the electrons, and $\beta_e = \omega/u_0$.

Equation (8) is a fourth-degree equation in terms of β and thus defines four possible x waves: two decaying waves in the medium, not associated with the beam, and two space-charge waves. The propagation factor of the space-charge waves in the beam is defined by the expression

$$\beta = \beta_e \pm \frac{\beta_p}{\sqrt{\epsilon_c}}, \quad (9)$$

where

$$\epsilon'_c = \epsilon' \left[1 - \frac{j\sigma}{\omega\epsilon} - \frac{B_L}{\omega\epsilon} \right], \quad \epsilon' = \frac{\epsilon}{\epsilon_0}; \quad (10)$$

ϵ'_c is the equivalent complex permittivity which includes the conductance and susceptance of the medium.

Substituting Equation (10) in Equation (9) it is seen that in the general case, with inductive-resistive reaction on the part of the medium, amplification of the space-charge waves in the beam is possible, that is,

$$\beta = \beta_e \pm \beta_p(p + jq) \frac{1}{\sqrt{\epsilon'}}, \quad (11)$$

where

$$p = \frac{\left\{ \frac{1}{2} \left[\left(1 - \frac{B_L}{\omega\epsilon} \right)^2 + \left(\frac{\sigma}{\omega\epsilon} \right)^2 \right]^{1/2} + \frac{1}{2} \left(1 - \frac{B_L}{\omega\epsilon} \right) \right\}^{1/2}}{\left[\left(1 - \frac{B_L}{\omega\epsilon} \right)^2 + \left(\frac{\sigma}{\omega\epsilon} \right)^2 \right]^{1/2}}; \quad (12)$$

$$q = \frac{\left\{ \frac{1}{2} \left[\left(1 - \frac{B_L}{\omega\epsilon} \right)^2 + \left(\frac{\sigma}{\omega\epsilon} \right)^2 \right]^{1/2} - \frac{1}{2} \left(1 - \frac{B_L}{\omega\epsilon} \right) \right\}^{1/2}}{\left[\left(1 - \frac{B_L}{\omega\epsilon} \right)^2 + \left(\frac{\sigma}{\omega\epsilon} \right)^2 \right]^{1/2}}. \quad (13)$$

The maximum values of p and q occur at magnitudes of $B_L/\omega\epsilon$ defined by the expressions

$$\left(\frac{B_L}{\omega\epsilon} \right)_{p_{\max}} = 1 - \frac{\sigma/\omega\epsilon}{\sqrt{3}}, \quad (14)$$

$$\left(\frac{B_L}{\omega\epsilon} \right)_{q_{\max}} = 1 + \frac{\sigma/\omega\epsilon}{\sqrt{3}}. \quad (15)$$

In which case

$$p_{\max} = q_{\max} \simeq \frac{0.81}{(\sigma'/\omega\epsilon)^{1/2}}. \quad (16)$$

Figures 1 and ω show families of curves for $p = f(B_L/\omega\epsilon)$ and $q = f(B_L/\omega\epsilon)$ with various values of the parameter $\sigma/\omega\epsilon$. In the case of small losses in the medium there occurs considerable amplification of the space-charge waves in the beam with $B_L/\omega\epsilon$ slightly larger than unity. With $0 < B_L/\omega\epsilon < 1$ significant amplification is possible only with considerable losses in the medium, while with $\sigma/\omega\epsilon = 0$ in this region of values of $B_L/\omega\epsilon$ there is no amplification. An increase in the losses in the medium decreases the maximum value of amplification (with $\sigma/\omega\epsilon \rightarrow 0$ $q \rightarrow \infty$), but the region of the parameters $B_L/\omega\epsilon$ in which amplification is possible is enlarged. The parameter p (determining the change in the characteristic plasma frequency) with $\sigma/\omega\epsilon = 0$ lies within the limits $1 \leq p \leq \infty$, when $0 \leq B_L/\omega\epsilon \leq 1$ and $p = 0$ with $B_L/\omega\epsilon \geq 1$. Thus, in the absence of losses in the medium, amplification of space-charge waves in the beam is possible for $B_L/\omega\epsilon \geq 1$ with an undisturbed velocity of the amplified wave $v_{ph} = u_0$. Losses lead to a decrease in p_{\max} and an increase in the region of frequencies within which p is other than zero.

For a physical picture of the interaction in the presence of inductive reaction of the medium let us examine the special case of the absence of loss ($\sigma/\omega\epsilon = 0$). In this case the induced charge in the medium is defined, in accordance with Equation (6), by the expression

$$\rho_c = -\rho_d \left[1 - \frac{1}{B_L/\omega\epsilon} \right]^{-1}. \quad (17)$$

Figure 3 shows the variation of $\rho_c/\rho_d = f(B_L/\omega\epsilon)$. With $B_L/\omega\epsilon \leq 1$ there is induced in the inductive medium a charge of the same sign as in the beam, and with $B_L/\omega\epsilon \rightarrow 1$ $\rho_c \rightarrow \infty$. In addition, the electric field intensity E'_c resulting from the induced charge is

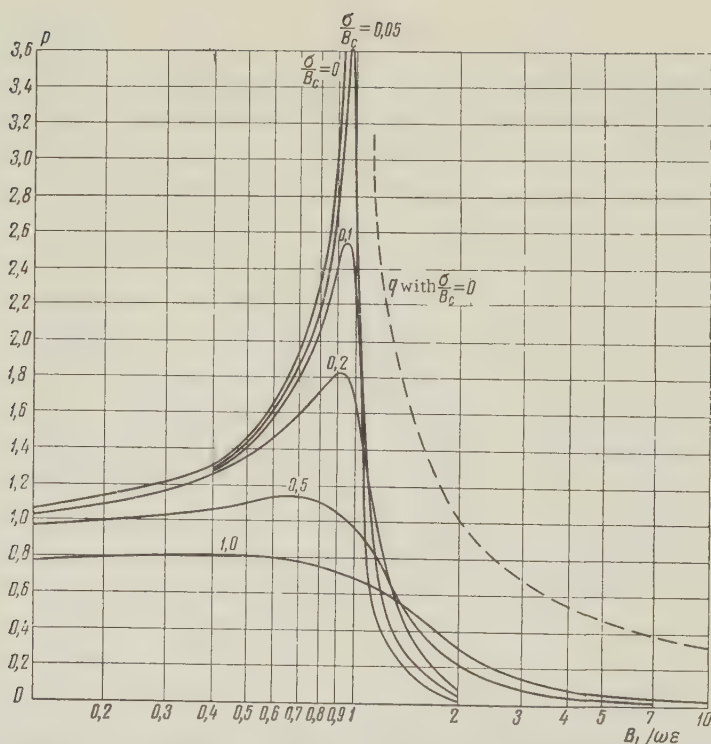


Figure 1 Parameter p (change in plasma frequency) as a function of the characteristics of the medium $B_L/\omega\epsilon$ and σ/B_c ($B_c = \omega\epsilon$)

added in phase with the space-charge intensity E_n in the beam, with the result that there is an increase in the longitudinal repulsion of electrons and amplification of the space-charge waves proves impossible. With $B_L/\omega\epsilon \geq 1$ the sign of the charge induced in the medium changes. The space-charge forces in the beam and the forces created by the induced charge E'_c are opposite in phase (see Figure 4). With $|\rho_c| \geq |\rho_d|$ the resulting intensity of the electric field acting in the beam leads to the acceleration of retarded electrons and to the retardation of accelerated electrons relative to the center of the bunch, resulting in an improvement of electron bunching in the beam and ensuring amplification of the space-charge waves. Upon fulfillment of the condition $B_L/\omega_0\epsilon = 1$ resonance exists in the medium: the displacement currents and the induction currents in the medium become equal in value. In this case the charge induced in the medium approaches ∞ and changes sign in the vicinity of $\omega = \omega_0$.

The amplification of space-charge waves in the passage of the beam through a system with inductive reaction is determined by the improvement of bunching of the electrons in the beam along the transit path and, as a result, an increase in the induced current in the medium. Power is not transmitted along the system. In this connection, in order to excite waves in the beam it is necessary to use a buncher (of the TWT delay-section type or a resonator), causing, in the general case, velocity and density modulation of the beam. For amplification in the inductive drift a device similar to the buncher is used, wherein the amplified variable component of current in the beam excites electromagnetic oscillations.

Let us conduct an approximate examination of the nonlinear effect of bunching in the inductive drift. In connection with the amplification of the variable components of current and velocity in the beam nonlinear effects are more intense here than in a drift with ideally conducting walls.

If the amplitudes of the variable components of current and electron velocity at the input into the inductive drift are i_{10} and u_{10} respectively, then the variable component of velocity z of an electron flying into the drift at moment t_0 is defined by the expression

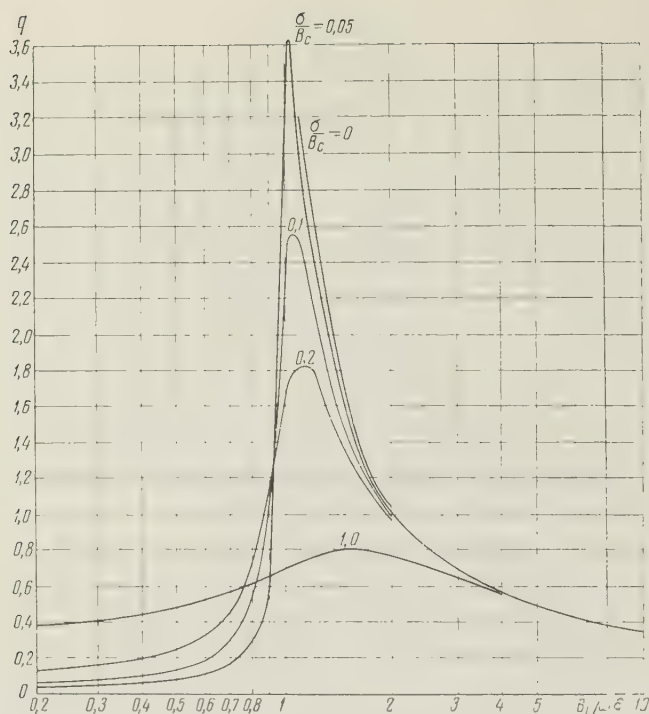


Figure 2. Gain parameter q as a function of the characteristics of the medium $B_L / \omega \epsilon$ and σ / B_C ($B_C = \omega \epsilon$).

$$u_1(z) = u_{10} \operatorname{ch} k_p q z \sin \omega t_0 + i_{10} \frac{u_0 q \omega_p}{\omega I_0} \operatorname{sh} k_p q z \cos \omega t_0. \quad (18)$$

Electron transit time within the drift ($z = s$) is

$$t_1 - t_0 = \int_0^s \frac{dz}{u_0 + u_{10} \operatorname{ch} k_p q z \sin \omega t_0 + i_{10} \frac{u_0 q \omega_p}{\omega I_0} \operatorname{sh} k_p q z \cos \omega t_0}. \quad (19)$$

Upon satisfying the conditions

$$\frac{u_{10}}{u_0} \operatorname{ch} k_p q s \ll 1, \quad \frac{i_{10}}{I_0} \frac{q \omega_p}{\omega} \operatorname{sh} k_p q s \ll 1 \quad (20)$$

we obtain from Equation (19)

$$\Theta = \omega(t_1 - t_0) \simeq \omega \tau - X \sin(\omega t_0 - \Psi), \quad (21)$$

where $\tau = s/u_0$,

$$\Psi = \arctg \frac{\frac{i_{10}}{I_0} (\operatorname{ch} k_p q s - 1)}{\frac{u_{10} \omega}{u_0 \omega_p q} \operatorname{sh} k_p q s}, \quad (22)$$

$$X = \left\{ \left(\frac{u_{10} \omega}{u_0 \omega_p q} \operatorname{sh} k_p q s \right)^2 + \left[\frac{i_{10}}{I_0} (\operatorname{ch} k_p q s - 1) \right]^2 \right\}^{1/2}, \quad (23)$$

where X is the bunching parameter.

As is seen from Equation (22), the bunching parameter for electrons within the inductive

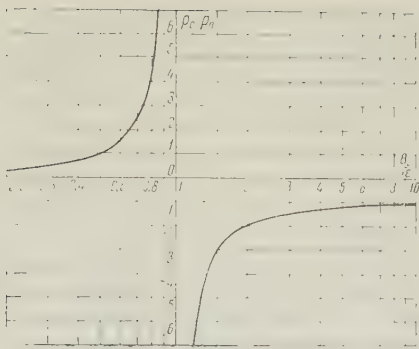


Figure 3. Charge ρ_c induced in the medium as a function of the characteristics $B_L/\omega\epsilon$ of the medium in the absence of losses ($\sigma/\omega\epsilon = 0$).

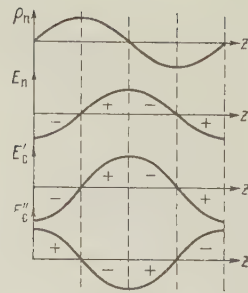


Figure 4. Characteristics of the longitudinal distribution of charge density within the beam ρ_d , space-charge field strength E_d and induced-charge field strength E'_c (with $B_L/\omega\epsilon < 1$) and E''_c (with $B_L/\omega\epsilon > 1$).

drift increases rapidly with an increase in the transit length s (almost exponentially for large values of $k_p q s$). The amplitude of the first harmonic of the current within the drift tube at a distance s from the buncher is determined from the expression

$$i_{1m} = i_{10} J_0(X) + 2I_0 \cos \psi J_1(X), \quad (24)$$

$J_0(X)$ and $J_1(X)$ are the zeroth-order and first-order Bessel functions.

The curves in Figure 5 show $i_{1m}/2I_0 = f(s)$ calculated from Equations (22) and (23) at various values of $\omega_p q/\omega$. In these calculations it was assumed that $u_{10}/u_0 = 0.01$ and $i_{10}/I_0 = 0.05$. The region of application of these ratios is limited by the extremely rigid conditions (20). The bunching characteristics at large values of $k_p q s$ may easily be determined in the special case where $i_{10} = 0$ and the excitation within the drift is achieved by velocity modulation of the electrons. In this case the electron transit angle in the drift

$$\omega(t_1 - t_0) = \omega \int_0^s \frac{dz}{u_0 + u_1 \operatorname{ch} k_p q z \sin \omega t_0} \quad (25)$$

and with $m = u_{10}/u_0 \ll 1$, which is always attainable in practice,

$$\Theta = \omega \tau - \frac{\omega}{\omega_p q} \ln \left[\frac{1 + m \operatorname{ch} k_p q s \sin \omega t_0}{1 + m \sin \omega t_0} \right]. \quad (26)$$

The curves in Figure 6 show $\omega t_1 = f(\omega t_0)$, determined from Equation (26) with $m \operatorname{ch} k_p q s = 0.05, 0.1$ and 0.15 and $\omega_p q/\omega = 0.72$. The broken curves show $\omega t_1 = f(\omega t_0)$ as calculated from Equation (21) for $m \operatorname{ch} k_p q s = 0.15$. Comparison of these curves confirms the possibility of using Equations (21)–(23) for approximate evaluation of the first harmonic of the current.

From an examination of the curves for $i_{1m} = f(s)$ it follows that, all other conditions being equal, the nature of the frequency dependence of the gain $K = i_1(f, s)/i_{10}$ may be varied by the choice of the length of drift with inductive reaction. Thus, with a short length of drift, when large values of $k_p q$ correspond to the optimum bunching condition, the gain decreases rapidly with an increase in $B_L/\omega\epsilon$, that is, with a decrease in frequency relative to resonance ($\omega = \omega_0$). At such a drift length, when the bunching optimum is attained at a frequency $\omega_1 < \omega_0$, the gain characteristic will have its maximum at a frequency close to ω_1 . In the region of frequencies $\omega_0 > \omega > \omega_1$, due to overbunching ($X > X_{om}$) amplification within the inductive drift will decrease. All other conditions being equal, with a constant length of inductive drift the nature of the frequency dependence of the gain is determined

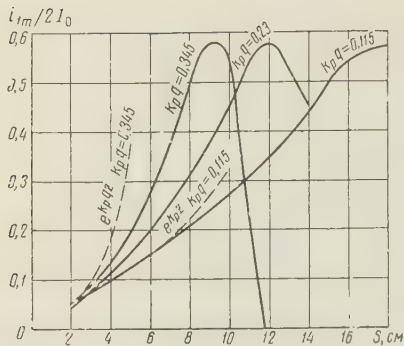
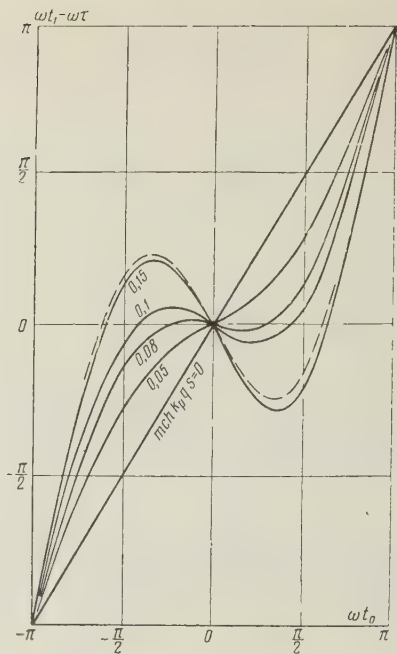


Figure 5. Change in the first harmonic of beam current along a drift with inductive reaction at various linear gains k_{pq} (in nepers).

Figure 6. Dependence of the electron arrival phase ωt_1 on the departure phase ωt_0 with various values of m ch ($k_{pq}S$).



by $u_{10}(f)$ and $i_{10}(f)$, that is the choice of operating conditions and the characteristics of the exciter section. For example, if u_{10} and i_{10} increase with a decrease in frequency, then we may obtain sufficiently small variations in gain within the inductive drift over a wide range of frequencies.

REFERENCES

1. J.R. Pierce, *Traveling-Wave Tubes*, Izd. Sovetskoye radio, 1952.
2. C.K. Birdsall, J.R. Whinnery, *J. Appl. Phys.*, 1953, 24, 3, 314.
3. E. Nalos, *IRE Trans.*, 1958, ED-5, 3, 161.
4. C.K. Birdsall et al., *Proc. IRE*, 1953, 41, 7, 865.

Submitted to the editors 22 April 1960.

CALCULATION OF A MAGNETIC PERIODIC FOCUSING DEVICE FOR TWT

A. L. Igritskiy

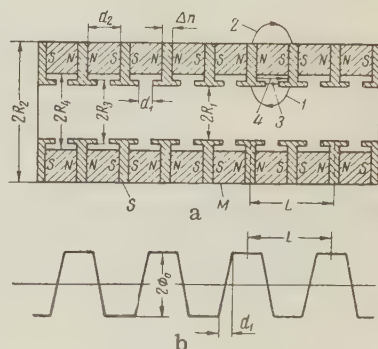
This paper discusses a method of calculating a magnetic system for periodic focusing of an electron beam in a TWT which permits consideration of the influence of the material and the dimensions of the pole pieces as well as of a finite number of magnets on the induction of the resulting periodic field. The method also permits consideration of the influence of disturbances of the periodicity and the ends of the device on the induction.

The calculation was carried out by the method of successive approximations. The method described ensures better agreement between calculated and experimental results.

INTRODUCTION

The use of periodic magnetic fields for focusing of the electron beam in traveling-wave tubes (TWT) permits a substantial reduction in the size and weight of the focusing device and also eliminates the need for external sources of power. The diagram of a device for periodic focusing is shown in Figure 1. Permanent ring magnets M (individually magnetized in the axial direction) are so stacked that the facing polarities of adjacent rings are identical. Between the permanent magnets there are pole pieces S of a material possessing a high saturation induction (e.g., Armco steel). The pole pieces contribute to field symmetry and greater field induction along the axis.

Figure 1. a - Longitudinal cross-section of magnetic focusing device: M - magnet; S - pole pieces; 1, 2, 3, 4, paths for fluxes ψ_1 , ψ_2 , ψ_3 and ψ_4 , respectively; b - distribution of magnetic potential over radius $R = R_1$.



Design calculations of a periodic focusing device are intended to determine the optimum shape and dimensions of the principal elements of the magnetic system: the permanent magnets and pole pieces. A method of calculation of periodic focusing devices was discussed in References 1 and 2.

As investigations have shown, calculations performed by this method yield values for induction which are 20-40% larger than experimental values. This difference in values is explained chiefly by the fact that the method of calculation described in References 1 and 2 does not consider the effect of the reluctance of the pole pieces on the value of induction, although this effect may be quite significant.

1. CALCULATION OF THE MAGNETIC SYSTEM IN THE FIRST APPROXIMATION

In calculation of the magnetic system in the first approximation we disregard the reluctance of the pole pieces. It is seen from Figure 1 that the pole pieces form a series of cylinders of identical diameter wherein the magnetic potentials of adjacent cylinders are opposite and equal to $\pm\Phi_0$ while in the gap between them at radius $r = R_1$ the potential varies linearly (see Figure 1b). The distribution of magnetic potential in the inner region between pole pieces in this case may be determined by solution of the Laplace equation; it has the form [1]

$$\Phi(z; r) = \sum_{n=1}^{\infty} \frac{2\Phi_0}{\sigma_1(n\pi)^2} \frac{J_0\left(i \frac{2n\pi}{L} r\right)}{J_0\left(i \frac{2n\pi}{L} R_1\right)} (1 - \cos n\pi) \sin(n\pi\sigma_1) \sin\left(\frac{2n\pi}{L} z\right), \quad (1)$$

$$\sigma_1 = \frac{d_1}{L}. \quad (2)$$

In Equations (1) and (2) L is the period of the magnetic field, R_1 is the inner radius of the pole pieces, d_1 is the distance between pole pieces, and $J_0\left(\frac{2n\pi}{L}r\right)$ is a zeroth-order Bessel function of the first kind with imaginary argument.

The component of the magnetic field induction along the z -axis may be defined as the gradient of the magnetic potential and, in absolute value is

$$B_z = \mu_0 \frac{\partial \Phi}{\partial z} = \sum_{n=1}^{\infty} \frac{4\Phi_0 \mu_0 (1 - \cos n\pi)}{L J_0\left(i \frac{2n\pi}{L} R_1\right)} \frac{\sin(n\pi \sigma_1)}{n\pi \sigma_1} J_0\left(i \frac{2n\pi}{L} r\right) \cos\left(\frac{2n\pi}{L} z\right). \quad (3)$$

The magnetic field along the z -axis (with $r = 0$) will vary in accordance with

$$B_{z0} = \mu_0 \sum_{n=1}^{\infty} \frac{4\Phi_0 (1 - \cos n\pi)}{L J_0\left(i \frac{2n\pi}{L} R_1\right)} \frac{\sin(n\pi \sigma_1)}{n\pi \sigma_1} \cos\left(\frac{2n\pi}{L} z\right). \quad (4)$$

The total magnetic flux in the plane $z = 0$ between pole pieces in the inner region

$$\psi_1 = \int_0^{R_1} B_z 2\pi r dr = 2\Phi_0 P_1. \quad (5)$$

Here P_1 is the permeance for flux ψ_1 :

$$P_1 = \mu_0 \sum_{n=1}^{\infty} \frac{-i J_1\left(i \frac{2n\pi}{L} R_1\right)}{J_0\left(i \frac{2n\pi}{L} R_1\right)} \frac{\sin n\pi \sigma_1}{n\pi \sigma_1} \frac{2(1 - \cos n\pi)}{n} R_1; \quad (6)$$

$J_1\left(i \frac{2n\pi}{L} R_1\right)$ is a first-order Bessel function of first kind with imaginary argument.

The leakage flux existing beyond the limits of the ring magnets (with $R > R_2$) also has considerable effect on the value of the maximum attainable field along the axis.

The leakage flux between pole pieces in the outer region is calculated in the same manner as the effective flux in the inner region and for it we derive the expression [1]

$$\psi_2 = 2\Phi_0 P_2. \quad (7)$$

Here P_2 is the permeance for the leakage flux:

$$P_2 = \sum_{n=1}^{\infty} \frac{-H_1^{(1)}\left(i \frac{2n\pi}{L} R_2\right)}{i H_0^{(1)}\left(i \frac{2n\pi}{L} R_2\right)} \frac{\sin(n\pi \sigma_2)}{n\pi \sigma_2} \frac{2(1 - \cos n\pi)}{n} R_2, \quad (8)$$

$$\sigma_2 = \frac{d_2}{L}. \quad (9)$$

In Equations (8) and (9) R_2 is the outer radius of the magnet, d_2 is the thickness of a single magnet, $H_0^{(1)}$ and $H_1^{(1)}$ are Hankel functions of the first kind and of the zeroth and first orders, respectively.

In addition to the magnetic fluxes ψ_1 and ψ_2 in the focusing system there exist the magnetic flux ψ_3 between the end faces of the annular pole pieces and the flux ψ_4 in the gap between the ring of a pole piece and the inner surface of the magnet.

The paths of these fluxes are shown in Figure 1 by 3 and 4. In accordance with Figure 1 the permeances for fluxes ψ_3 and ψ_4 may be calculated from

$$P_3 = \mu_0 \frac{\pi (R_3^2 - R_1^2)}{d_1}, \quad (10)$$

$$P_4 = \mu_0 \frac{\pi (R_4^2 - R_3^2)}{d_2}. \quad (11)$$

Here R_3 is the outer radius of the annular pole piece, R_4 is the inner radius of the permanent ring magnet.

The total permeance between pole pieces

$$P_0 = \sum_{i=1}^4 P_i. \quad (12)$$

If the difference of magnetic potential between pole pieces is $2\Phi_0$, then each magnet will create a magnetic flux

$$\psi_0 = 2\Phi_0 P_0. \quad (13)$$

The induction in the permanent magnet is

$$B_m = \frac{2\Phi_0 P_0}{\pi (R_2^2 - R_4^2)}. \quad (14)$$

Let us assume that the magnetic field strength is constant over the length of the magnet. Then from the definition of the field strength as the gradient of the potential we can write

$$H_m d_2 = 2\Phi_0, \quad (15)$$

where H_m is the field strength in the permanent magnet.

From Equations (14) and (15) we obtain the relationship between induction and field strength in the permanent magnet:

$$\frac{B_m}{H_m} = \frac{d_2 P_0}{\pi (R_2^2 - R_4^2)}. \quad (16)$$

On the other hand, B_m and H_m are related by the demagnetization curve $B = f(H)$. Figure 2 shows the demagnetization curve for anisotropic ferroxdur. If all the dimensions of the magnetic system are given, we can calculate the right hand side of Equation (16). Extending a straight line from the coordinate origin to the curve

$$\alpha = \arctg \frac{d_2 P_0}{\pi (R_2^2 - R_4^2)}, \quad (17)$$

we find the operating point of the permanent magnet (see Figure 2). Knowing the values of H_m and B_m at the operating point, we can ascertain how close the operating conditions of the magnets are to optimum conditions. Equations (15) and (4) permit plotting the distribution of the magnetic field along the axis of the periodic focusing device.

As the result of Calculation of the magnetic system in the first approximation we also determine the fluxes ψ_1 , ψ_2 , ψ_3 and ψ_4 , which will be used in calculation of the magnetic circuit in the second approximation.

2. CALCULATION OF THE MAGNETIC SYSTEM IN THE SECOND APPROXIMATION

In a similar calculation of the magnetic circuit let us consider the reluctance of the material of the pole pieces. Figure 3 shows the distribution of magnetic fluxes in the pole piece of a periodic device. It is not difficult to conclude from the figure that there must exist a certain separating radius R_p . The lines of force which proceed from the magnet at $R > R_p$ will form the leakage flux

$$\psi_2 = B_m \pi (R_2 - R_p^2), \quad (18)$$

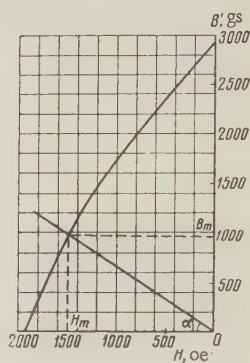


Figure 2. Demagnetization curve for anisotropic ferroxdur $B = f(H)$.

By calculating the value of B_R for several values of R and by using the magnetization curve we can plot the curve of $H_R = f_{ba}(R)$ for portion ba

$$U_{ba} = \int_{R_4}^{R_p} H_R dR = \int_{R_4}^{R_p} f_{ba}(R) dR. \quad (26)$$

Portion ad. At portion ad there flows an internal flux, the value of which varies from $\psi_B = \psi_1 + \psi_3 + \psi_4$ at $R = R_4$ to $\psi_1 + \psi_3$ at $R = R_3$. Through the section of radius R there flows a flux ψ_B minus the flux branching into the gap between pole pieces along path ψ_4 at the portion between R_4 and R (see Figure 3):

$$\psi_R = \psi_B - B\pi(R_4^2 - R^2). \quad (27)$$

Here B is the induction in the gap between pole pieces along the path ψ_4 :

$$B = \mu_0 \frac{2\Phi_0}{d_2} = \mu_0 H_m. \quad (28)$$

In the absolute system of units $\mu_0 = 1$ and the induction in the gap along path ψ_4 is numerically equal to the field strength in the magnet.

The induction in the pole piece section of radius R

$$B_R = \frac{\psi_B - H_m \pi (R_4^2 - R^2)}{\pi R \Delta p}. \quad (29)$$

By calculating the values of B_R for several values of R and using the magnetization curve we can plot $H_R = f_{ad}(R)$ for portion ad. The drop in magnetic potential at portion ad

$$U_{ad} = \int_{R_3}^{R_4} f_{ad}(R) (dR). \quad (30)$$

Portion de. At portion de there flows through the pole piece a flux which varies in value from $\psi_1 + \psi_3$ at point d to ψ_3 at the end face of the pole piece (point e). Through the circumferential cross-section at a distance z from the central plane of the gap there pass flux ψ_3 and the part of flux ψ_1 which leaves the inner surface of the pole piece cylinder and proceeds to a point with coordinate z :

$$\psi_z = \psi_3 + \int_0^z B_r(R_1; z) 2\pi R_1 dz. \quad (31)$$

Here $B_r(R_1; z)$ is the radial component of induction at the inner surface of the pole piece cylinder. It can be determined as the gradient of magnetic potential from Equation (1); in absolute value it will be

$$\begin{aligned} B_r(r; z) &= \mu_0 \frac{\partial \Phi}{\partial r} = \\ &= \mu_0 \sum_{n=1}^{\infty} \frac{-iJ_1\left(i\frac{2n\pi}{L}r\right) \sin n\pi\tau_1}{J_0\left(i\frac{2n\pi}{L}R_1\right) \frac{\sin n\pi\tau_1}{n\pi\sigma_1}} \frac{4\Phi_0(1 - \cos n\pi)}{L} \sin\left(\frac{2n\pi}{L}z\right). \end{aligned} \quad (32)$$

Substituting B_r into Equation (31) and integrating, we obtain

$$\begin{aligned} \psi_z &= \psi_3 + \\ &+ \mu_0 \sum_{n=1}^{\infty} \frac{-iJ_1\left(i\frac{2n\pi}{L}R_1\right) \sin n\pi\tau_1}{J_0\left(i\frac{2n\pi}{L}R_1\right) \frac{\sin n\pi\tau_1}{n\pi\sigma_1}} \frac{4\Phi_0(1 - \cos n\pi)}{n} R_1 \left[1 - \cos\left(\frac{2n\pi}{L}z\right)\right]. \end{aligned} \quad (33)$$

The induction in the pole piece cross-section with coordinate z is

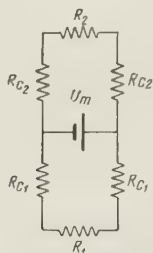
$$B_z = \frac{\psi_z}{\pi(R_4^2 - R_1^2)}. \quad (34)$$

Calculating the values of B_z for several values of z and using the magnetization curve for the pole piece material, we can plot the curve of $H_z = f_{de}(z)$ for portion de

$$U_{de} = \int_{\frac{d_1}{2}}^{\frac{L}{4}} H_z dz = \int_{\frac{d_1}{2}}^{\frac{L}{4}} f_{de}(z) dz. \quad (35)$$

If the drop in magnetic potential at portion de is considerable, this will lead to a change in the distribution of magnetic potential at radius $R = R_1$ (see Figure 1b) and it may prove convenient to derive a new expression for $\Phi(z; r)$ instead of Equation (1), which was derived for a trapezoidal potential distribution with $R = R_1$.

For an individual magnet of the periodic system we can provide the equivalent conversion circuit [3, 4] shown in Figure 4. Here R_1 takes into account the air resistance along the path of internal flux:



$$R_1 = \frac{2\Phi_0}{\psi_B} = \frac{1}{P_1 + P_3 + P_4}. \quad (36)$$

R_{C1} represents the (magnetic) resistance of the path of internal flux in steel. This resistance we shall relate to ψ_B and it will be

Figure 4. Equivalent conversion diagram of an individual magnet in the periodic focusing device.

$$R_{C1} = \frac{U_{ba} + U_{ad} + U_{de}}{\psi_B}. \quad (37)$$

Here U_{ba} , U_{ad} and U_{de} are the drops in magnetic potential at the corresponding portions, the values of which were defined above.

R_2 is the resistance in air to the leakage flux ψ_2 :

$$R_2 = \frac{2\Phi_0}{\psi_2} = \frac{1}{P_2}. \quad (38)$$

Resistance R_{C2} represents the resistance to the leakage flux ψ_2 in steel:

$$R_{C2} = \frac{U_{bc}}{\psi_2}. \quad (39)$$

Using the equivalent circuit, let us determine the total resistance of the magnetic circuit

$$R_\Sigma = \frac{R_{e1}R_{e2}}{R_{e1} + R_{e2}}. \quad (40)$$

Here $R_{e1} = R_1 + 2R_{C1}$ is the equivalent resistance for the internal magnetic flux and R_{e2} is the same for the leakage flux.

On the basis of Ohm's magnetic law the flux created by the magnet

$$\psi = B_m \pi (R_2^2 - R_4^2) = \frac{H_m d_2}{R_\Sigma}. \quad (41)$$

From Equation (41) we obtain the ratio between the induction B_m and the field strength H_m in the magnet in the second approximation:

$$\frac{B_m}{H_m} = \frac{d_2}{\pi (R_2^2 - R_4^2) R_\Sigma} = \text{tg } \alpha_e. \quad (42)$$

On the other hand, B_m and H_m are related by the demagnetization curve $B = f(H)$ shown in Figure 2. On the plot of $B = f(H)$ we extend a straight line through the coordinate origin at an angle of

$$\alpha_e = \arctg \frac{d_2}{\pi (R_2^2 - R_3^2) R_\Sigma}, \quad (43)$$

and find the operating point of the permanent magnet in the second approximation.

The mmf created by the magnet:

$$U_m = H_m d_2. \quad (44)$$

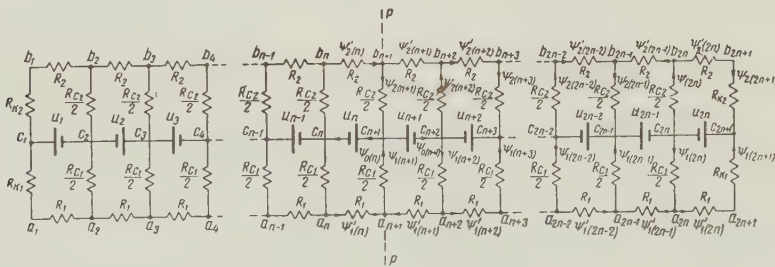
The induction of the magnetic field at the axis can be determined in the second approximation from Equation (4) if, instead of the difference of potentials between pole pieces $2\Phi_0$, we substitute the drop in magnetic potential at the resistance R_1 , which, as is easily seen, is

$$U_{R_1} = \frac{H_m d_2}{R_1 + 2R_{C_1}} R_1. \quad (45)$$

If the calculated value of B_{z0} in the second approximation differs but little from the value of B_{z0} obtained in the first approximation, then the calculation can be considered concluded. If there is significant departure in these values, the calculation must be repeated with the new values of mmf and fluxes obtained in the second approximation.

3. INFLUENCE OF THE ENDS OF THE FOCUSING DEVICE

In the case of calculating the reluctance of the pole pieces the magnetic circuits of the individual permanent magnets prove to be interrelated and for the entire periodic focusing system we may present the equivalent conversion diagram shown in Figure 5. Calculations from this conversion diagram permit evaluating the mutual influence of the magnets and also permit considering the influence of the ends of the focusing system on the distribution of the field within it.



The magnetic fluxes are

$$\psi'_{1(n)} = \psi'_{1(n+1)}; \quad \psi'_{2(n)} = \psi'_{2(n+1)}; \quad \psi_{0(n)} = \psi_{0(n+1)}.$$

Thus it suffices to perform the calculations for only half of the system, for example, to the right of the plane of symmetry PP. The currents and mmf's to the left are found from the conditions of symmetry. Calculation from the circuit shown in Figure 5 is conveniently performed by the method of successive approximations. On the first approximation let us consider that the mmf's of the permanent magnets and the resistances of the flux paths for iron are identical and known. The values of the mmf's and resistances are determined from Equations (44), (37) and (39). It is required to determine the magnetic fluxes flowing along the various sections of the magnetic circuit.

For the lower contour with number $n + 1$ and subsequent contours, on the basis of Kirchoff's second law [4], we can write the equations

$$\begin{aligned} \psi_{1(n+1)} \frac{R_{c1}}{2} + \psi'_{1(n+1)} R_1 + \psi_{1(n+2)} \frac{R_{c1}}{2} &= U, \\ \psi_{1(n+2)} \frac{R_{c1}}{2} + \psi'_{1(n+2)} R_1 + \psi_{1(n+3)} \frac{R_{c1}}{2} &= U, \\ \\ \psi_{1(n+k)} \frac{R_{c1}}{2} + \psi'_{1(n+k)} R_1 + \psi_{1(n+k+1)} \frac{R_{c1}}{2} &= U, \\ \\ \psi_{1(2n)} \frac{R_{c1}}{2} + \psi'_{1(2n)} R_1 + \psi_{1(2n+1)} R_{k1} &= U. \end{aligned} \quad (46)$$

Applying Kirchoff's first law to nodes a_{n+1} , a_{n+1} , etc, we obtain

$$\begin{aligned}
 \psi'_{1(n+1)} &= \psi'_{1(n)} = \frac{\psi_{1(n+1)}}{2}, \\
 \psi'_{1(n+2)} &= \psi_{1(n+2)} - \psi'_{1(n+1)} = \psi_{1(n+2)} - \frac{1}{2} \psi_{1(n+1)}, \\
 &\dots \dots \dots \\
 \psi'_{1(n+k)} &= \psi_{1(n+k)} - \psi'_{1(n+k-1)} = \\
 &= \psi_{1(n+k)} - \psi_{1(n+k-1)} + \dots - \psi_{1(n+3)} + \psi_{1(n+2)} - \frac{1}{2} \psi_{1(n+1)}, \\
 &\dots \dots \dots \\
 \psi'_{1(2n)} &= \psi_{1(2n+1)} = \psi_{1(2n)} - \psi'_{1(2n-1)} = \\
 &= \psi_{1(2n)} - \psi_{1(2n-1)} + \dots - \psi_{1(n+3)} + \psi_{1(n+2)} - \frac{1}{2} \psi_{1(n+1)}.
 \end{aligned}
 \tag{47}$$

The last two equations are applicable with k and n even.

Substituting the fluxes from Equation (47) into Equation (46), we have

$$\begin{aligned} & \frac{1}{2} \psi_{1(n+1)}(R_{c1} + R_1) + \psi_{1(n+2)} \frac{R_{c1}}{2} = U, \\ & \psi_{1(n+2)} \left(\frac{R_{c1}}{2} + R_1 \right) - \frac{1}{2} \psi_{1(n+1)} R_1 + \psi_{1(n+3)} \frac{R_{c1}}{2} = U, \\ & \dots\dots\dots \\ & \psi_{1(n+k)} \left(\frac{R_{c1}}{2} + R_1 \right) - \psi_{1(n+k-1)} R_1 + \dots + \psi_{1(n+2)} R_1 - \\ & \quad - \frac{1}{2} \psi_{1(n+1)} R_1 + \psi_{1(n+k+1)} \frac{R_{c1}}{2} = U, \\ & \dots\dots\dots \\ & \psi_{1(2n)} \left(\frac{R_{c1}}{2} + R_1 + R_k \right) - \psi_{1(2n-1)} (R_1 + R_k) + \\ & \quad + \dots + \psi_{1(n-2)} (R_1 + R_k) - \frac{1}{2} \psi_{1(n+1)} (R_1 + R_k) = U. \end{aligned} \tag{48}$$

Thus, we have obtained a system of n equations with n unknown fluxes $\psi_{1(n+1)}$, $\psi_{1(n+2)}, \dots, \psi_{1(2n)}$.

The solution of this system of equations by the usual method by means of determinants is accompanied by considerable difficulty in calculation. However, the solution is extremely simplified by the use of the following approximate method. Usually the resistance of the pole piece material $R_{c1}/2$ is considerably lower than the resistance R_1 for flux for air and the fluxes $\psi_{1(n+1)}$, $\psi_{1(n+2)}$ etc differ little from one another. Hence, in the first equation of system (48) we can with sufficient accuracy disregard the second expression. Flux $\psi_{1(n+1)}$ in the first approximation will be

$$\psi_{1(n+1)I} = \frac{2U}{R_{c1} + R_1}. \quad (49)$$

Substituting the resulting value of $\psi_{1(n+1)}$ into the second equation of the system and disregarding the second expression therein, we find $\psi_{1(n+2)}$, etc. Thus, we can find all the fluxes successively. If it is desirable to determine these fluxes with greater accuracy, the resulting values for the fluxes in the first approximation must be used in determining the fluxes in the second approximation. Thus, flux $\psi_{1(n+1)}$ in the second approximation is defined by the expression

$$\psi_{1(n+1)II} = \frac{2 \left[U - \psi_{1(n+2)I} \frac{R_{c1}}{2} \right]}{R_{c1} + R_1}. \quad (50)$$

The other fluxes can be determined in a similar fashion.

We have discussed a method of determining the internal fluxes ψ_1 of a magnetic system. Completely similar equations can be derived for the leakage fluxes ψ_2 . For this purpose Kirchoff's second law must be applied to the upper contours with numbers $n+1$ and following. Kirchoff's first law must be applied to nodes b_{n+1} , b_{n+2} , etc. The resulting system of equations is solved in the same manner as described above.

Having determined fluxes $\psi_{1(n+k)}$ and $\psi_{2(n+k)}$, the fluxes in the remaining sections of the magnetic circuit can be determined without particular difficulty. The operating points and mmf's of individual magnets of the periodic system can be determined. If the values of flux and mmf obtained differ but little from the values determined in Section 2, this calculation may be considered final. If there is significant difference in the values, another approximation will be required.

CONCLUSION

This paper has presented a method of calculation of a magnetic system for periodic focusing of the electron beam in a TWT. Its chief difference from the hitherto used Chang method lies in the consideration of the reluctance of the pole pieces. The method described permits the calculation of the influence of the pole piece material and dimensions on the induction of the resulting periodic field. It permits calculating the influence of disturbances of the periodicity and ends of the device on the value of induction and ensures better coincidence between the results of calculation and experiment.

REFERENCES

1. K.K.N. Chang, Optimum design of periodic magnet structures for electron beam focusing, RCA Rev., 1955, 16, 1, 65.
2. F. Sterzer, W.W. Siekanowicz, The design of periodic permanent magnets for focusing of electron beams, RCA Rev., 1957, 18, 1, 39.
3. M.A. Babikov, Electrical Devices, Part I, GEI, 1951.
4. A.Ya. Buylov, Principles of Construction of Electrical Devices, GEI, 1946.
5. Roters, Electromagnetic Mechanisms, GEI, 1949.

Submitted to the editors 29 February 1960

SOME EFFECTS ACCOMPANYING DETECTION IN A GASEOUS DISCHARGE

G.D. Lobov and V.I. Yeremeyev

This paper presents an experimental comparison of the characteristics of the detection of microwave oscillations in a gaseous discharge with the characteristics of enhanced glow of the same discharge. Both effects are investigated for helium and neon gas-discharge tubes at various powers and discharge currents.

INTRODUCTION

The action of an electromagnetic microwave field on a gaseous discharge has already been described in the literature. It has been shown, for example, that in applying microwaves to a gas-discharge tube both the value of the discharge current [1] and the glow level of the gas [2, 3] vary. In the work described here an increase in the excitation glow of atoms and a decrease in the recombination glow in the microwave field were investigated in connection with the process of detection in a glow discharge.

EXPERIMENT

Experimental investigation of the effects of excitation glow and quenching of recombination glow was performed with the setup shown in the block diagram in Figure 1.

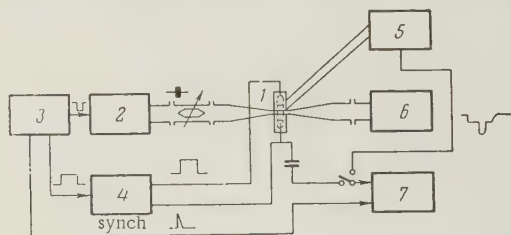


Figure 1. Block diagram of measurement setup.

1 - Gas-Discharge detector; 2 - microwave generator; 3 - synchronizer; 4 - power supply for gas-discharge detector; 5 - photomultiplier with power supply; 6 - wattmeter; 7 - oscillograph.

Microwave pulses of duration $\tau_p = 10 \mu\text{sec}$ were applied to a gas-discharge detector through a waveguide. For the investigation of the influence of microwave power on the various sections of discharge, the waveguide was narrowed to a slot at the point of insertion of the gas-discharge detector. This permitted application of the microwaves to narrow sections of the gaseous discharge. The gas-discharge detector was a glass tube with plane metal electrodes between which a glow discharge was excited. The tube was filled with an inert gas.

The glow discharge in the detector could arise from both a d-c voltage and the application of special ignition pulses of duration $\tau_{ig} = 75 \mu\text{sec}$. In such ignition of the gas-discharge detector microwave pulses could be applied at different moments, including during the afterglow of the gas.

The brighteners of the glow discharge and all changes in this glow under the influence of the microwave field were detected by means of a photomultiplier through a slit in the narrow wall of the waveguide. The voltages from the photomultiplier and the load resistor of the gas-discharge detector were applied to an oscillograph.

2. RESULTS

In all cases during the action of sufficient microwave power on the discharge the intensity of the glow at the point of action of the pulse was increased. This increase in glow differed for neon and helium. With a decrease in microwave power the glow dropped to zero and then there appeared a pulse of reverse polarity (that is, a quenching pulse caused by a decrease in recombination glow).

Figure 2 shows oscillograms of the voltage at the photomultiplier output, obtained during the action of microwaves on the discharge created by d-c voltage in helium. The oscillograms are shown in order of increasing microwave power. At low powers there is seen the quenching pulse of the recombination glow at the point of action of the microwave pulse (Figure 2a).

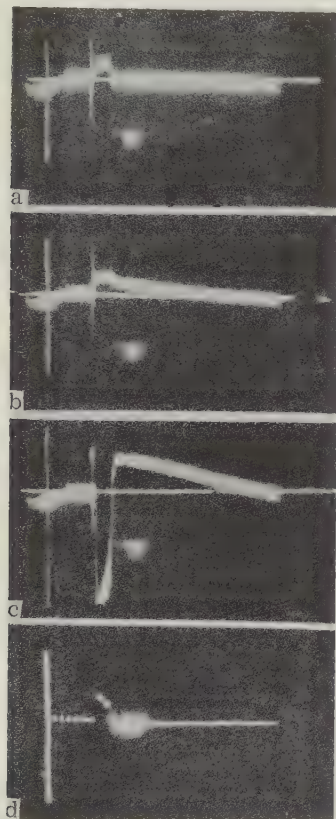


Figure 2. Oscillograms of pulses of extinction and increase of glow for a discharge by d-c voltage in helium, $p = 9$ mm Hg, $\tau_p = 20$ μ sec. Photographed at different microwave powers. a - $P/P_m = 0.01$; b - $P/P_m = 0.05$; c - $P/P_m = 1$; d - gas-discharge detector pulse.

With an increase in microwave power the trailing edge of the pulse gradually decays. This may be explained by the fact that the electrons, having received energy during the pulse, retain it for a while after the pulse, which decreases the probability of recombination. The greater the applied microwave power, the greater the decrease in recombination. With a further increase in microwave power there appears at the site of the quenching pulse a pulse of intensified excitation glow accompanied by a distinct overshoot after the pulse.

In a discharge by d-c voltage no quenching pulse is observed in a neon-filled tube.

For more distinct observation of the recombination glow, pulse ignition of the gas-discharge detector was employed. Figure 3 shows the oscillograms for a neon tube.

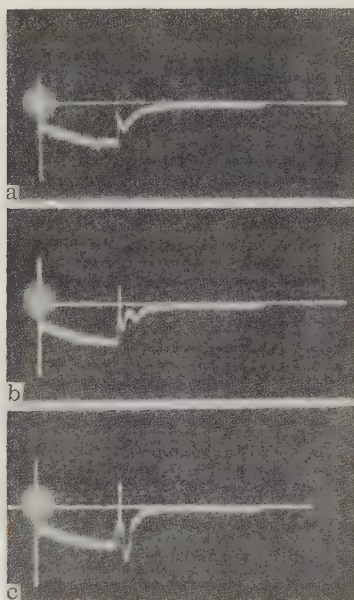


Figure 3. Oscillograms of extinction pulses and an increased glow pulse in neon afterglow for different microwave powers. $p = 21$ mm Hg; $\tau_p = 10$ μ sec, $\tau_{ig} = 75$ μ sec. a - $P = 0$ w; b - $P/P_m = 0.08$; c - $P/P_m = 1$.

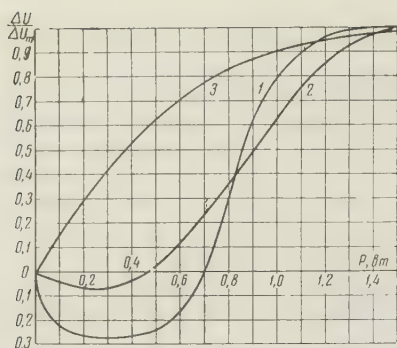


Figure 4. Dependence of pulse magnitudes at the output of the photomultiplier (ΔU_{pn}) and the gas-discharge detector (ΔU_d) on power for a gas-discharge tube with helium; $p = 9$ mm Hg, $\tau_p = 10 \mu\text{sec}$, $\tau_{ig} = 75 \mu\text{sec}$:

1 - $\Delta U_{pn}(P)$, microwave pulse occurs during afterglow; 2 - $\Delta U_{pn}(P)$ microwave pulse occurs during the action of the ignition pulse; 3 - $\Delta U_d(P)$.

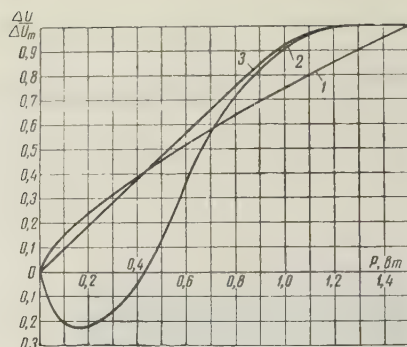


Figure 5. Dependence of pulse magnitudes at the photomultiplier output (ΔU_{pn}) and the gas-discharge detector output (ΔU_d) on microwave power applied to a neon gas-discharge tube; $p = 21$ mm Hg, $\tau_p = 10 \mu\text{sec}$, $\tau_{ig} = 75 \mu\text{sec}$:

1 - $\Delta U_{pn}(P)$ for afterglow; 2 - $U\Delta_{pn}(P)$ for ignition; 3 - $\Delta U_d(P)$.

Oscillograms for helium tubes were similar and hence are not presented here. Figures 4 and 5 show the dependence of the magnitude of the pulses from the photomultiplier and the gas-discharge detector on applied power for helium and neon tubes. From these curves and the oscillograms it is evident that the quenching effect is observed in the helium tubes and in the neon tubes, but is stronger in the helium tubes. Thus, for tubes filled with neon the quenching pulse is observed only in the afterglow (Figure 5, Curves 1 and 2). In the helium-filled tube extinction of the recombination glow occurs both upon the action of the microwave field during the ignition pulse and after it (Figure 4).

Several curves were plotted for clarification of the relation between detection and enhanced glow. Thus, Figure 6 shows the distribution of the glow and the dependence of pulse magnitudes from the gas-discharge detector and photomultiplier on the length of discharge in helium. The maximum of the detected pulse coincides with the excitation maximum; both of these maxima occur in the negative glow. Since the discharge spread beyond the electrodes, the pulses were also observed beyond the discharge gap. Curves were plotted for neon-filled tubes and were similar to those shown in Figure 6.

Figure 7 shows the distribution of the extinction and the excitation of glow along the length of the discharge in a tube filled with helium. The maximum of the quenching effect and the maximum of the increase in glow are slightly out of coincidence.

The dependence of the alternating voltage at the gas-discharge detector and the photomultiplier output on discharge current is shown in Figure 8. The detector voltage has a distinct maximum as the discharge current is varied.

3. EVALUATION OF RESULTS

The experimental relationships derived reveal a similarity in the phenomena of detection and variation in glow. A qualitative explanation of these results may be obtained on the assumption that the microwave power changes the electron energies.

If we assume (as in Reference 1) that the current depends on the electron energies,

$$I_0 = A Q^n,$$

where I_0 is the value of the discharge current and Q is the electron energy, then an electron energy increment of ΔQ will cause a current increment

$$\Delta I = n I_0 \frac{\Delta Q}{Q}. \quad (1)$$

On the other hand, electrons possessing energy Q experience, according to Reference 4, the total number of collisions leading to excitation

$$\left(\frac{dN}{dt}\right)_{\text{exc}} = B \left(\frac{Q}{Q_{\text{exc}}}\right)^{\frac{5}{2}} e^{-\left(\frac{Q_{\text{exc}}}{Q}\right)^2}.$$

With a change in energy by ΔQ the increase in the number of exciting collisions will be

$$\delta \left(\frac{dN}{dt}\right)_{\text{exc}} = \left(\frac{dN}{dt}\right)_{\text{exc}} \left[\frac{5}{2} + 2 \left(\frac{Q_{\text{exc}}}{Q}\right)^2 \right] \frac{\Delta Q}{Q}. \quad (2)$$

If it is considered that the intensity of glow Φ is proportional to the number of exciting collisions, then a change in the number of collisions will lead to a change in glow by $\Delta \Phi$, which, in accordance with Equation (2), is

$$\Delta \Phi_{\text{exc}} = \Phi_0 \left[\frac{5}{2} + 2 \left(\frac{Q_{\text{exc}}}{Q}\right)^2 \right] \frac{\Delta Q}{Q}. \quad (3)$$

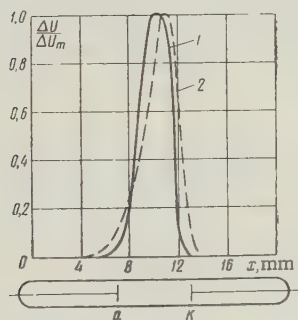


Figure 7. Distribution of quenching (1) and excitation (2) of the glow over the length of the discharge; $p = 9$ mm Hg, Helium, $\tau_p = 20 \mu\text{sec}$, $P/P_m = 0.05$.

Hence it follows that an increase in energy leads to an increase in the excitation glow of atoms, wherein in the first approximation the increment of glow is proportional to the total excitation glow. This fact has been observed in actual operation (Figure 6).

The decrease in glow upon the application of low power is most easily explained if it is considered that the recombination coefficient decreases with increasing electron energy in proportion to $Q^{-1/2}$ [4]. Since the recombination coefficient indicates the number of electrons recombined per second, it characterizes the degree to which the

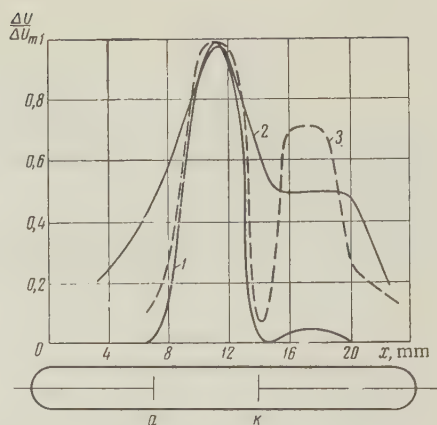


Figure 6. Pulse magnitude at output of the photomultiplier and the gas-discharge detector as functions of discharge length: $p = 9$ mm Hg, helium, $\tau_p = 10 \mu\text{sec}$, $\tau_{ig} = 75 \mu\text{sec}$:

1 - $\Delta U_{pn}(x)$; 2 - $\Delta U_d(x)$; 3 - distribution of glow $U_g(x)$.

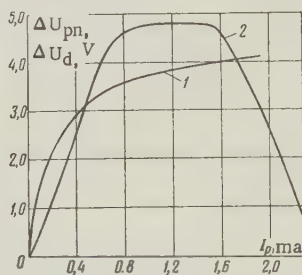


Figure 8. Dependence of pulse values at the output of the photomultiplier (1) and the output of the gas-discharge detector (2) on discharge current; $p = 9$ mm Hg, Helium, $\tau_p = 10 \mu\text{sec}$, $\tau_{ig} = 75 \mu\text{sec}$, $P/P_m = 1$.

glow depends on recombination. Hence it follows that an increase in energy by ΔQ will lead to a decrease in glow by

$$\Delta\Phi_{\text{rec}} = -\frac{1}{2}\Phi_{\text{rec}}\frac{\Delta Q}{Q}. \quad (4)$$

Due to the fact that the regions of intense recombination and excitation do not coincide, the functions $\Delta\Phi_{\text{rec}}$ and $\Delta\Phi_{\text{exc}}$ have their maxima at different points of the gas discharge (Figure 7).

As follows from Equations (1), (3) and (4), the magnitudes of the effects in detection as well as in enhanced glow are identically dependent on the increment in electron energy and, consequently, to this degree must be similar in nature.

CONCLUSIONS

The curves and oscillograms show that detection in a gaseous discharge is accompanied by a change in glow. The nature of the dependences of the change in glow on various parameters (power, discharge current and site of the radiation) coincides qualitatively with the analogous dependences for detection. To a certain degree this relationship explains the physical nature of the operation of the gas-discharge detector. Actually, while detection is not always accompanied by a change in recombination glow, the recombination coefficient, as this work has shown, does vary. Since a change in this coefficient may lead to a change in the number of free electrons, then the discharge current may also change to the same extent.

A change in the number of excited electrons upon application of the microwave field may also serve as a source of change in the discharge current (due to secondary processes). On the other hand, similarly, a change in the number of collisions leading to excitation also changes the number of collisions accompanied by ionization. This phenomenon will also cause a change in discharge current.

REFERENCES

1. G.D. Lobov, Gas-discharge detector of microwaves, *Radiotekhnika i elektronika*, 1961, 6, 1, 117.
2. L. Goldstein, J. Anderson, Quenching of afterglow in gaseous discharge plasmas by low power microwaves, *Phys. Rev.*, 1953, 90, 486.
3. J.M. Anderson, Quenching of the negative glow by microwaves in cold-cathode gaseous discharges, *Phys. Rev.*, 1957, 108, 898.
4. A. Engel', *Ionized Gases*, Fizmatgiz, 1959.

Moscow Power Engineering Institute
Chair of Radio Engineering Theory

Submitted to the editors 6 June 1960

EMISSION OF HOT ELECTRONS FROM p-n JUNCTIONS IN SiC CRYSTALS

M.I. Elinson, G.V. Stepanov and V.I. Pokalyakin

Emission of hot electrons from p-n junctions in crystals of SiC is investigated as a function of the magnitude of the cut-off voltage at the junction U_j and temperature T . The existence of a high current density and a nonuniform distribution of emission from the surface is confirmed. The magnitude of the emission current i_e is determined by the current through the junction i_{rev} . The ratio γ has a maximum at the specific voltage U_j .

INTRODUCTION

SiC crystals are suitable objects for the study of emission of hot electrons, since $\epsilon_i > \chi$ [ϵ_i - is the threshold energy of impact ionization, χ - is the electron affinity). For hexagonal SiC $\epsilon_i \approx 4.3$ eV, $\chi = 4$ eV. Moreover SiC crystals are rather inert chemically which allows us to expect stability in their surface properties.

In Reference 2 for the first time the emission of electrons from natural p-n junctions in SiC was investigated. A number of interesting features of the emission were found, in particular, a high current density $j_e > 1$ amp/cm² and the law $i_e \sim i_{rev}^k$, where i_e is the emission current; i_{rev} is the reverse current through the junction. There exist several opinions on the nature of i_{rev} [3, 4, 5], important for the understanding of the mechanism of emission. Mainly References 4 and 5 give a concept of incompletely ionized donor centers in SiC and their ionization which seems more plausible from a comparison with the artificial considerations expounded in Reference 3 concerning p-n junctions. Unfortunately, in Reference 2 we did not investigate the most important variations of i_e with electric field in the junction and with temperature. Also the nature of emission centers is not clear.

1. OBJECTS AND METHOD OF INVESTIGATION

Just as in Reference 2 we investigated natural p-n junctions which form during the growth of SiC crystals by the method of sublimation. The existence of junctions was established from the well-known method of electroluminescence and by the volt-ampere characteristic. The crystals were cut from nodules, polished and selected by the number and brightness of luminous points upon switching on the cut-off voltage. The surface treatment consisted of the removal of SiO₂ dielectric films and outgassing.

The experimental apparatus and measuring circuit are given in Figure 1. Crystal K was clamped to a thick molybdenum lead Ct with resilient tungsten points O. The crystal temperature was measured by thermocouple T. The apparatus also contained a collector grid C connected to a willemite screen F onto a semitransparent platinum base. This base allowed the comparison of spots on the screen with the position of luminous points on the crystal.

On to the crystal one could apply a static as well as a pulsed voltage. The latter enabled us to maintain a constant crystal temperature even at high reverse voltages U_j .

The anode voltage was adjusted over wide limits, and the emission current was measured by an electrometer with sensitivity $\sim 10^{-13}$ amp.

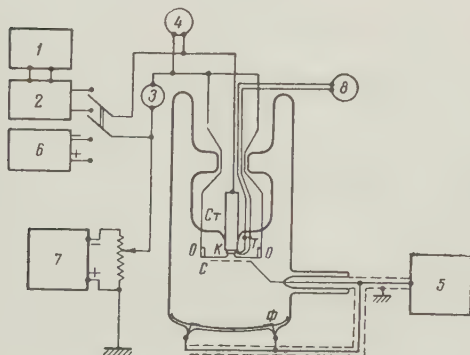


Figure 1. Experimental apparatus and circuit.

K - SiC crystal; O - tungsten clamping points; Ct - thin molybdenum rod for enhancement of heat exchange with the crystal; C - collector anode grid; F - fluorescent screen on a metallic semitransparent base; 1 - pulse oscillator 26I; 2 - pulse amplifier; 3 - device for measuring stationary or average current; 4 - device for measuring d-c or pulse voltage; 5 - vacuum tube electrometer; 6 - d-c current source; 7 - anode voltage source; 8 - temperature measuring instrument.

2. RESULTS OF INVESTIGATION AND DISCUSSION

In a number of crystals, emission currents from 10^{-12} to 10^{-6} amp were obtained. As a rule, emission occurred in due course, which we produced with a continuous sampling

of current and with an increase of crystal temperature (~ 400 deg C.). After approximately eight hours of warmup, the emission attained a stationary value and became exceedingly stable. The formation of emission is connected perhaps with the cleanliness of the crystal surface.

The junction voltage U_j necessary for emission varies within wide limits for different specimens. Comparison of the emission picture on a luminescent screen with the pattern of luminous points on the crystal shows that the luminous points are emission centers. The large number and the great intensity of the luminous points are not definitely associated with a large emission current. The junction depth distribution near the emitting surface is the determining value. The number of emitting centers increase with an increase of U_j . The linear dimensions of the emitting centers lie within the limits 1 to 10 μ . The current density calculated from the total area of emitting centers is 1-10 amp/cm². Hence, the most important result of Reference 2, concerning the high emission current density is confirmed.

Figure 2 gives the relations $i_{rev} = f(U_j)$ and $i_e = f(U_j)$ for one emitter. The coarse structure of the characteristic for the reverse current (curves 1 and 2) consists of a section of rapid increase $i_{rev}(AB)$ and a section of slower variation of $i_{rev}(BC)$ (see also Figure 3). Diminution in the growth of i_{rev} is somewhat unusual for a cut-off voltage.

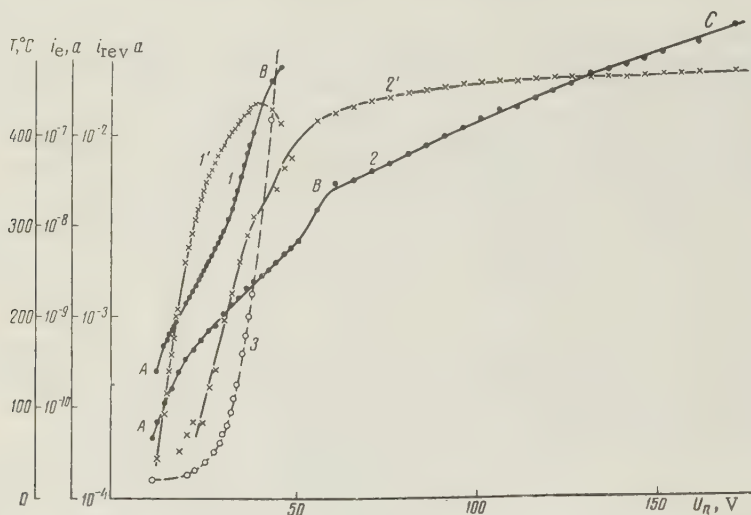


Figure 2. Comparison of static and pulse characteristics of a junction.

1 - $i_{rev}(U_j)$ for the static condition; 1' - $i_e(U_j)$ for the static condition; 2 - $i_{rev}(U_j)$ for the pulsed condition, $T = 20^\circ \text{C}$; 2' - $i_e(U_j)$ for the pulsed condition, $T = 20^\circ \text{C}$; 3 - layer temperature during operation in the static condition.

Curve 1 corresponds to a static regime in which the crystal temperature changed greatly with an increase of U_j (the temperature change is described in Curve 3). Curve 2 corresponds to pulsed operation ($\tau_j = \mu\text{sec}$), when over the entire range of U_j the crystal is maintained at room temperature.

The corresponding characteristic curves of the emission current i_e are given in curves 1' and 2'.

The region AB corresponds to a rapid rise in i_e . In region BC, the rise in i_e is greatly diminished (more than i_{rev}). For static operation, due to a continuous temperature increase, a smaller diminution in the growth of i_e is observed. Only at $T > 400$ deg C there occurs a drop in i_e , possibly connected with an increase of electron scattering by lattice vibrations.

Figure 3 gives the pulse characteristics $i_{rev}(U_j)$ and $i_e(U_j)$ taken for two crystal

temperatures $T = 20^{\circ}\text{C}$ and $T = 75^{\circ}\text{C}$.

On the basis of Figures 2 and 3 one may infer that the influence of temperature on hot electron emission under these conditions is weaker and of a different nature than that assumed in the theory [1], which applied to Ge and Si type crystals, especially in the region BC.

Hence in the system studied, the increase of electron scattering with a rise in temperature T is compensated either by an increase in the number of electrons, or possibly, by a change in the electric field in the junction upon varying T .

The behavior of the characteristic also shows that the voltage at which virtual "saturation" of the emission current is achieved is very far from threshold.

The following established fact is especially interesting. If i_e is plotted as a function of i_{rev} (see Figure 4), all experimental characteristics, independent of junction voltage U_j and temperature T , within the limits of experimental errors, agree with one another, i.e., the experimental value of i_{rev} always corresponds entirely to a specific value of i_e .

This may be interpreted so that the electrical field at the junction E is very large and depends weakly on U_j (for example, $E \sim \sqrt{U_j}$). Therefore the acceleration of electrons at the junction is always sufficient and is virtually independent of U_j and T . The emission current is larger, the more the carriers at the junction.

Figure 4 also illustrates the ratio $\gamma = i_e/i_{rev}$, common for all characteristics. This ratio has a maximum in the break region of the characteristic $i_{rev}(U_j)$ (region of points B).

Consequently, γ depends on the behavior of i_{rev} and can be explained by exterior causes, for example by junction emission in the condition of space charge.

Evidently, in the region BC a new strong scattering mechanism comes into play, and the number of electrons capable of being emitted increases considerably more slowly than the total number of current carriers.

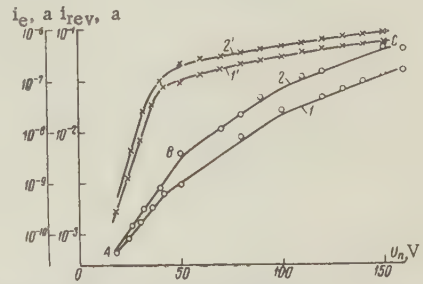


Figure 3. Pulse characteristics $i_{rev}(U_j)$ and $i_e(U_j)$ at different temperatures:

1 - $i_{rev}(U_j)$ at 20°C ; 1' - $i_e(U_j)$ at 20°C ; 2 - $i_{rev}(U_j)$ at 75°C ; 2' - $i_e(U_j)$ at 75°C .

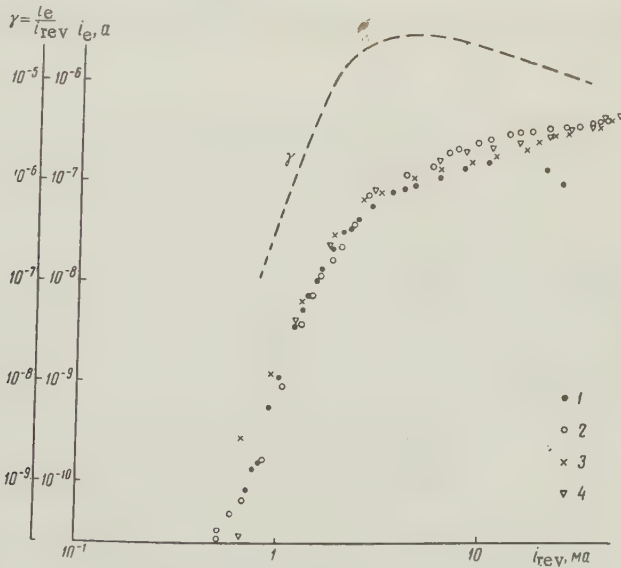


Figure 4. Characteristics $i_e(i_{rev})$; $\gamma(i_{rev})$
1 - static; 2, 3 - pulse, at 20°C ; 4 - pulse, at 75°C .

The relation between i_{rev} and i_e also appears clearly in characteristic curves taken at liquid nitrogen temperature (see Figure 5; the curves refer to another crystal). It was established that a sudden increase of currents leading to the descending character of the relation is connected with heating of the crystal.

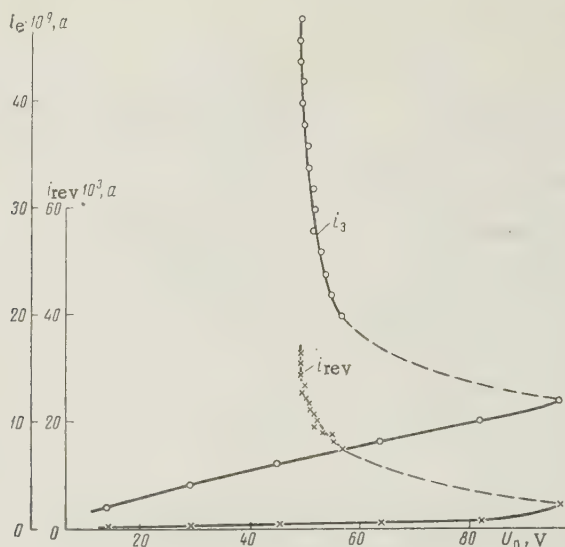


Figure 5. Static characteristics $i_{\text{rev}}(U_j)$ and $i_e(U_j)$ at -180°C .

The rigid relation i_e/i_{rev} indicates that the back current passes in considerable part through those places of the junction, which are emission centers. It indicates the constancy of emitting surface properties.

The maximum value is $\gamma \approx 10^{-4}$, i.e., very small. This is connected, evidently, with electron energy losses in the strongly alloyed substrate n-th layer in which the electric field is insignificant.

The increased efficiency of this type of emission is one of the chief problems for further study.

CONCLUSIONS

In the study of emission of hot electrons from p-n junctions in SiC the following facts were established.

1. The emission current density, in agreement with Reference 2, is rather large $j_e = 1 - 10 \text{ amp/cm}^2$.
2. The emission is uniformly distributed over the surface. Luminescent points on the crystal surface are emission centers.
3. The temperature variation of emission is less and has a different character in comparison with the theoretical relation for homogeneous Ge and Si-type semiconductors, which is connected with factors compensating for an increase in electron scattering by lattice vibrations when temperature is increased.
4. The ratio γ of emission current to reverse current through the junction is very small ($\sim 10^{-4}$), which possibly is connected with electron scattering near the emitting surface and has a maximum at a specific voltage U_j .

The decrease in γ after the maximum is related to a new strong scattering mechanism.

5. The emission current is closely connected with the reverse current through the junction, independent of the temperature T and voltage U_j at which the experimental value of reverse current is obtained. This fact is explained by the existence of a strong electric field at the junction, always ensuring sufficient electron acceleration.

The authors express their gratitude to V.B. Sandomirskii for useful discussion, and also to N.V. Sumin and A.M. Fadeeva for assistance in the work.

REFERENCES

1. Sh.M. Kogan and V.B. Sandomirskii, *Fiz. Tverd. Tela* 2, 10, 2570 (1960).
2. L. Patrick, W.J. Choyke, *Phys. Rev. Let.*, 1959, 2, 2, 48.
3. L. Patrick, *J. Appl. Phys.*, 1957, 28, 765.
4. R. Goffaux, *J. phys. et radium*, 1960, 21, 2, 94.
5. E.T. Karlamova and G.F. Kholuyanov, *Fiz. Tverd. Tela* 2, 13, 426 (1960).

Submitted to the editors 7 September 1960

FIELD EMISSION FROM TUNGSTEN SINGLE CRYSTALS IN THE PREBREAKDOWN CURRENT REGION

G.N. Fursei

Results are given for an investigation of field emission in the "prebreakdown" period; the design of the apparatus and the experimental technique are described.

INTRODUCTION

It is well-known that at a current density above a certain critical value, time-stationary field emission passes into a "vacuum arc"-avalanche process, by melting the emitter [1]. This process leads to the destruction of the cathode, and hence it is interesting to study phenomena which lead to a "vacuum breakdown" and its associated phenomena. Papers by Dyke [1, 2, 3, 4, 5], Elinson [6, 7, 8], as well as Shuppe and Gorman [9] in recent years have been devoted to this subject.

These authors used a pulse technique which allowed the removal from the emitter of considerably larger current densities than in a stationary regime, and the observation of a number of phenomena, which would be difficult to study in the stationary regime.

It was found that the arc is preceded by a number of phenomena, which are anomalous in comparison with the usual flow of the process; namely: 1) spontaneous growth of current during the action of a rectangular voltage pulse; 2) deviation from the form of the volt-ampere characteristic predicted by the Fowler-Nordheim theory which is well substantiated at smaller current densities; 3) the emergence of a bright ring surrounding the usual emission picture.

The goal of our investigation was to check the existence of these phenomena, as well as the results published in References 1-8 and in Reference 9 which show a discrepancy.

This discrepancy is that in References 1-8, the volt-ampere characteristic deviates from rectilinearity at small currents, but in Reference 9 - in the region of larger currents.

Assuming that the technique used by us possesses a somewhat greater resolving power and sensitivity, we hope to obtain additional information concerning the development of the process.

1. CONDITIONS AND METHOD OF INVESTIGATION

In order to carry out the investigation, equipment (Figure 1) consisting of the following

main units was assembled.

1. A generator of high-voltage rectangular pulses of duration $0.5 \mu\text{sec}$ at a repetition rate of 50-1000 pulses/sec. The generator could also give separate pulses of length $0.5-10 \mu\text{sec}$ with smooth amplitude regulation within the limits 0-20 kv and a rise time less than $0.1 \mu\text{sec}$. This generator included a type 26I pulse generator, which served as the source unit, and a modulator (M) of two tubes (GI-30 and GMI-83) which operated in a keyed mode of operation with partial capacitive discharge. The use of the 26I generator considerably simplified the assembly of the equipment because it was necessary merely to build a very simple modulator (Figure 2).

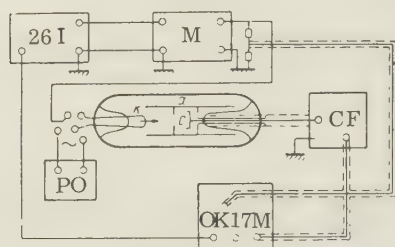


Figure 1. Block diagram of the equipment.

the collector of the experimental device in order to minimize stray capacitance and in that way decrease the input circuit time constant, and also to match the total load resistance with the input amplifier of oscilloscope OK17M (100 ohm). On the second pair of plates in the oscilloscope cathode ray tube we imposed the voltage pulse taken from a resistance divider.

2. Double-beam oscillograph OK17M with a photographic attachments.

3. Cathode follower (CF) with an attenuator at the input, which enabled us to change the input signal amplitude in steps up to a factor of 100.

4. Heating circuit for the point (PO).

A negative high-voltage pulse with an amplitude of several kilovolts and length two μsec was incident on cathode K of the experimental vacuum device (Figure 1). The pulse of field emission current, which originated was detected by collector C and proceeded through the cathode follower to the first amplifier of oscilloscope OK17M. The cathode follower was connected directly to

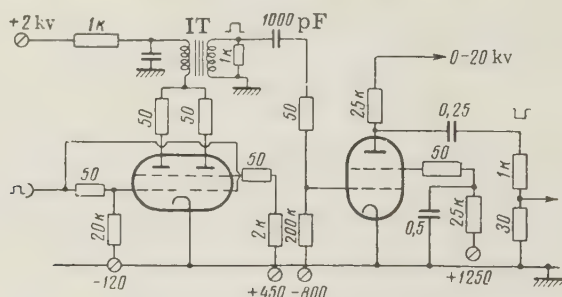


Figure 2. Schematic diagram of the modulator.

Thus on the oscilloscope screen one could simultaneously observe the voltage pulse and its corresponding current pulse at the collector.

Triggering of the equipment was carried out by an intensifier pulse from the OK17M oscilloscope. Synchronized operation of the equipment and oscilloscope sweep circuit was ensured by an appropriate selection of sweep time in the 26I oscillator and oscilloscope.

The shutter of the photographic apparatus was carefully opened and in one frame an exposure was made at three-four points.

The amplifier and plates of the oscilloscope tube had been previously calibrated by means of a GSS-6, and also by means of a square wave pulse generator and a pulse voltmeter. Both calibration curves agree within the limits of error of the equipment used for the calibration.

Measurements of the magnitude of current and voltage were made directly on the film by means of a comparator.

Voltage resolution was restricted by the width of the sweep trace on the film and was equal to 0.05 mm, which corresponds to a voltage of approximately 50 v. Current resolution was also restricted by the width of the trace, but since with an insignificant voltage variation the current changed substantially the current lines were virtually always "resolved".

Errors in the current and voltage measurements were determined chiefly by the class of device used to calibrate the oscillograph, and could be estimated as 5-7%.

Tungsten was chosen as the material whose emission and vacuum properties were to be most carefully studied.

A cathode in the form of a point was prepared by the standard method of automatic electrolytic etching which enabled us to manufacture the point to extremely precise values of radii. This emitter was placed in a device which was subsequently evacuated. The vacuum system was improved by spraying a titanium and molybdenum absorber and by evacuation with an Alpert manometer. The measurements were made at a pressure of the order of 10^{-9} mm Hg. Before recording each point, the point was heated in order to clean its surface. Surface purity was controlled through emission pictures which were observed at the anode slit.

2. DESIGN OF THE APPARATUS

The main difference in the technique described above was the use of screening in the vacuum device.

Great difficulty in the measurements is created by the existence of very strong pickup from the electrode onto which a high-voltage pulse is fed. To prevent this effect one may decrease the load resistance which makes it possible to obtain a time constant at which the pulse pickup is differentiated and timewise amounts to only an insignificant part of the useful signal (decrease of R), or via internal screening (decrease of C).

The second method described by us is simpler, because, by taking a signal with a comparatively large resistance (10 kohm), one could measure rather weak currents (10^{-5} amp), by using the wide-band amplifier of oscillograph OK17M without any alteration. In addition, without changing the gain, and also without leading to additional calibration, one could change the input resistance by steps which allows a 1000-fold decrease in the signal and a simultaneous improvement in the condition for the passage of the current pulse.

In this work we mainly used two types of vacuum devices (Figure 3). Their principal element was a screening cylinder.

In design 1 (Figure 3) the point is located inside the cylinder close to the aperture of the anode slit so that the emitted electrons could not impinge on clear glass

which would lead to its charging and to field distortion. The collector section with a flexible screen was connected to the input of the cathode follower. Design 3 was different from 1 in that the screening role was played by a metallic layer deposited on the glass.

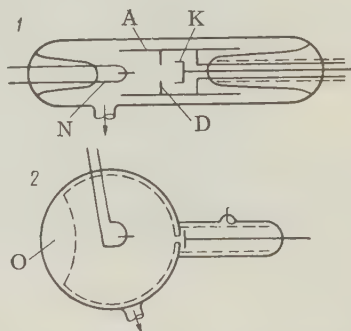


Figure 3. Design of experimental apparatus.

K - collector; A - anode cylinder; D - slit;
N - cathode; O - window for observation of
of emission pictures.

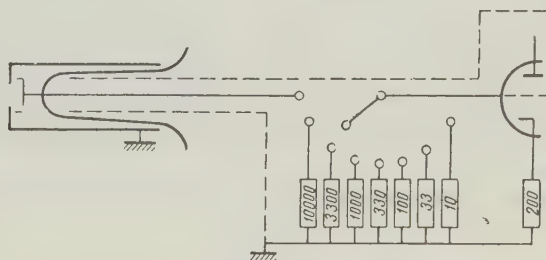


Figure 4. Collector section and cathode follower input.

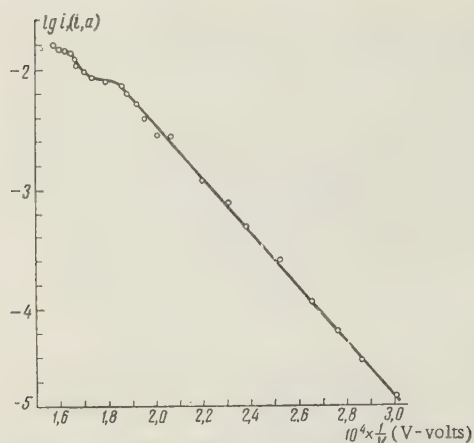


Figure 5. Volt-ampere characteristic of the field emission current during pulsed operation.

case. For this purpose we supplied the cathode with a two μ sec length voltage pulse at a repetition rate of 250 pulses/sec and measured galvanometrically the average total anode current and collector current. By a similar method we recorded up to 20 points and then determined the average value of the attenuation coefficient.

3. RESULTS

Control measurements were made on 15 emitters in vacuum devices of a typical type. These measurements were allowed us to establish the following.

1. At large current densities in all 15 cases, the volt-ampere characteristic deviates toward the region of smaller current value (Figure 5).

2. In the "nonlinear" part of the characteristics, periodic current oscillations occur. A similar type of undulating pulsation of thermionic emission was observed in the presence of strong fields [10]. In Reference 11 these differences are explained by quantum-mechanical barrier effects of the interference of electron waves reflected from the potential barrier. We are inclined to assume that a similar type of phenomena also occur in our case. By confining ourselves for the present to the assumption stated above, we hope to conduct a careful analysis of this phenomena in the next paper.

3. We observed a spontaneous rise in current with time (Figure 6), which was discovered at a current density of approximately 30-40% of the critical value which agreed with the results of References 1-5.

4. During a calculation of the emitter radii according to the approximate formula of Reference 12, we calculated critical current densities which amounted to from $3(10)^7$ to $6(10)^7$ amp/cm², which also is in good agreement with data obtained by other authors.

5. Immediately before the emergence of the arc we noticed a spontaneous decrease of current at constant voltage, which supposedly can be attributed to the effect of space charge (Figure 7).

6. As can be seen from the oscillogram of the moment of the emergence of an arc (Figure 8), the "arc process" is very short-lived and lasts less than 10^{-7} sec.

The author gives heartfelt thanks to I. L. Sokol'skaya for recommending this work and for reviewing the manuscript. The author thanks Y. U. V. Zubenko and G. P. Shcherbakov for practical advice related with high vacuum technology. It is a pleasant obligation to express my sincere gratitude to I. E. Isaev, Yu. V. Markov and the graduate A. A. Rogachev for a helpful conversation on the subject of technique.

The screening cylinder (Figure 4) served to shield the collector from pickup inside the device. Such a design allowed us to decrease the pickup approximately 1000 fold with a 5-15 fold reduction in the useful signal. Thus, the screen succeeded in almost completely eliminating the effect of the pickup, which resulted only in an insignificant distortion of the leading edge of the pulse. The point was situated at a distance of 5-10 mm from the aperture, so that it would be possible to project onto the collector the greater part of the emission picture.

At the maximum sensitivity ($R = 10$ kilohm) one could confidently measure signals corresponding to a collector current of the order 10^{-6} amp. The time constant of the circuit allowed us to handle pulses of useful signal with very little distortion. At other positions of the input attenuator, corresponding to resistances of 3.3, 1 kilohm, 333, 100, 33, 10 ohms, similar distortions are absent.

The decrease brought about by the screening cylinder was determined experimentally for each

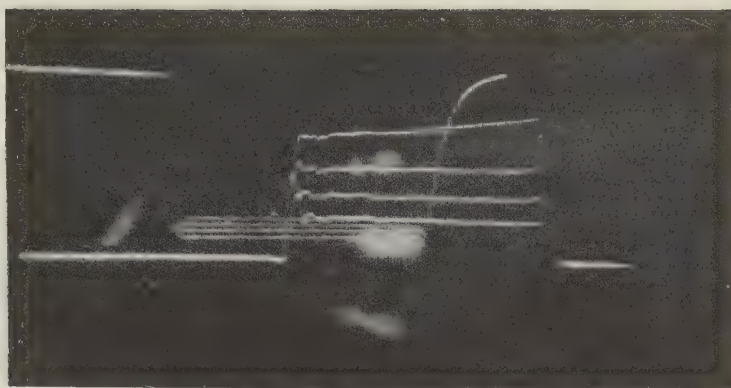


Figure 6. Emergence of spontaneous growth of current.

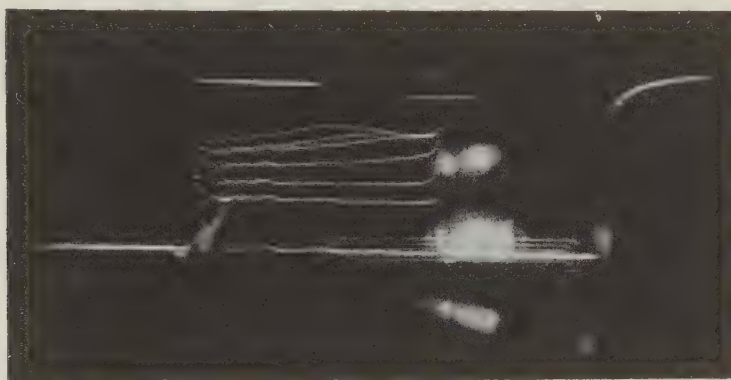


Figure 7. Moment of spontaneous growth of current, decrease of current and the formation of an arc.

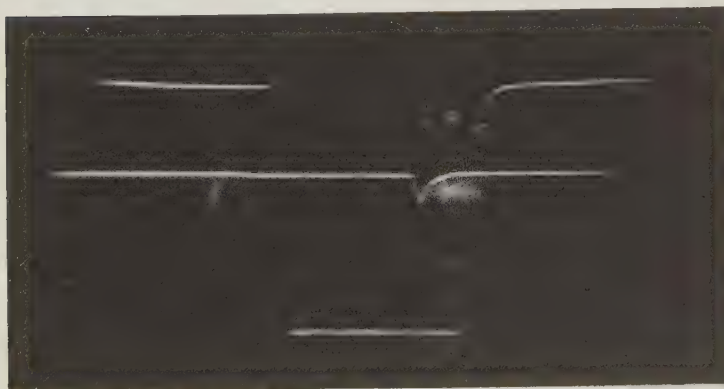


Figure 8. Oscillogram of the voltage pulse and current pulse at the moment of breakdown.

REFERENCES

1. W.P. Dyke, L.K. Trolan, *Phys. Rev.*, 1953, 89, 4, 799.
2. W.P. Dyke, J.K. Trolan, E.E. Martin, J.P. Barbour, *Phys. Rev.*, 1953, 91, 5, 1043.
3. W.P. Dyke, J.K. Trolan, E.E. Martin, *Phys. Rev.*, 1953, 91, 5, 1054.
4. J.P. Barbour, W.W. Dolan, J.K. Trolan, E.E. Martin, W.P. Dyke, *Phys. Rev.*, 1953, 92, 1, 45.
5. W.P. Dyke, J.P. Barbour, E.E. Martin, J.K. Trolan, *Phys. Rev.*, 1955, 99, 4, 1192.
6. M.I. Elinson, G.F. Vasil'yev, *Field Emission*, Fizmatgiz, 1958.
7. M.I. Elinson, G.F. Vasil'yev, *Radiotekhnika i Elektronika*, 1958, 3, 7, 945.
8. M.I. Elinson, V.A. Gor'kov, A.A. Yasnopol'skaya, G.A. Kudintseva, *Radiotekhnika i Elektronika*, 1960, 5, 8, 1318.
9. I.I. Gofman, O.D. Protopopov, G.N. Shupe, *Izv. AN UzSSR*, 1959, 6, 73; I.I. Gofman, G.N. Shupe, O.D. Protopopov, *Fizika Tvyerdovgo Tyeia*, 1960, 2, 6, 1323.
10. R.Z. E. Siefert, T.E. Phipps, *Phys. Rev.*, 1939, 56, 7, 663.
11. E. Guth, C.J. Mullin, *Phys. Rev.*, 1941, 59, 7, 575.
12. M. Drechsler, E. Henkel, *Z. angew. Phys.*, 1954, 6, 8, 341.

Submitted to the editors 1 July 1960

ON THE THEORY OF THE STATIC CHARACTERISTICS OF A PLANAR DIODE WITHOUT AN INCANDESCENT CATHODE TAKING INTO ACCOUNT THERMAL ELECTRON VELOCITY DISTRIBUTION

G. Ya. Myakishev

Expressions are obtained which determine the static characteristic of a planar diode without an incandescent cathode for different regimes. In contrast with known papers we take into account the velocity distribution of electrons. Conditions are analyzed for a transition from one regime to another.

INTRODUCTION

At the present time the static characteristic of a planar diode without an incandescent cathode is investigated in detail under the condition that the electron velocity distribution can be neglected [1]. Such an approximation, generally speaking, is permissible, because the mean thermal velocity is usually much less than the velocities electrons gain in an accelerating field.

However with the formation of a virtual cathode, a calculation of the electron velocity distribution, especially in the region near the virtual cathode, may in a number of cases have great significance, in particular, during a study of diode noise and during the investigation of the influence of high frequency perturbations on the electron flux and retardation by the electric field [2]. The virtual cathode brings about a sorting of electrons according to velocity.

It is interesting therefore to investigate the influence of the thermal distribution of electrons with respect to velocity on the static characteristic of a diode without an incandescent cathode. It is of interest also to clarify the limits of existence of different types of diode operating regimes.

We limit ourselves to a consideration of the most important case when the electron

beam accelerated by the grid falls into a retarding electric field between the grid and the reflector (reflex klystron). The expression obtained may be extended to the case of an accelerating field.

1. STATEMENT OF THE PROBLEM

Let the grid be located at the coordinate origin, and the reflector at a distance d from it. The electron velocity distribution and the grid plane, through which electrons pass in this diode, will be taken into consideration by means of the distribution function corresponding to a solution of the problem of a planar diode with an incandescent cathode* (for velocities $v > 0$). Electron scattering by the grid structure is neglected. We will also assume that the reflector completely absorbs electrons incident on it (no secondary emission exists).

Under these conditions the distribution function has the form

$$\begin{aligned} f(0, v) &= \Theta(v) N_0 \sqrt{\frac{m}{2\pi kT}} e^{-\frac{mv^2}{2kT} + \frac{e\varphi_1}{kT}} \quad \text{for } v > 0, \\ f(d, v) &= 0 \quad \text{for } v < 0, \end{aligned} \quad (1)$$

φ_1 is the grid- to -cathode potential difference; N_0 - is the electron concentration; $\Theta(v)$ - is a cut-off factor that takes into account the fact that the electron velocity at the grid cannot be less than specific values:

$$\Theta(v) = \begin{cases} 1 & \text{for } v \geq \sqrt{\frac{2e}{m} V_1}, \\ 0 & \text{for } v < \sqrt{\frac{2e}{m} V_1}; \end{cases} \quad (2)$$

$V_1 = \varphi_1 + \varphi_2$, where φ_0 is the depth of the potential minimum relative to the cathode in the cathode-grid space. In the absence of a virtual cathode in this space $V_1 = \varphi_1$.

The boundary conditions for the potential are chosen in the form

$$\varphi(0) = \varphi_1, \quad \varphi(d) = \varphi_2, \quad (3)$$

where φ_2 is the reflector potential relative to the cathode. In this case $\varphi_1 \geq \varphi_2$.

To solve the problem we use the system of equations adopted by A.A. Vlasov [4], by assuming that inter-electron collisions can be neglected.

As a result of the plane symmetry of the problem and its static nature, the equations for the distribution and potential functions have the form

$$\begin{aligned} v \frac{\partial f(x, v)}{\partial x} + \frac{e}{m} \frac{d\varphi(x)}{dx} \frac{\partial f(x, v)}{\partial v} &= 0, \\ \frac{d^2 \varphi(x)}{dx^2} &= 4\pi e \int_{-\infty}^{+\infty} f(x, v) dv. \end{aligned} \quad (4)$$

The sign of the electron charge is taken into account here

We introduce further dimensionless magnitudes:

$$\begin{aligned} x' &= x \sqrt{\frac{4\pi e^2 N_0}{kT}}, \\ t' &= t \sqrt{\frac{4\pi e^2 N_0}{m}}, \quad v' = v \sqrt{\frac{m}{kT}}, \\ \varphi' &= \varphi \frac{e}{kT}, \quad f' = f \sqrt{\frac{kT}{m}} \frac{1}{N_0}. \end{aligned} \quad (5)$$

*Such a function better represents the actual physical picture than the Maxwellian drift function discussed in Reference 3.

Omitting the primes in what follows (for simplicity of notation). The original system of equations and boundary conditions may be rewritten in terms of new dimensionless quantities in the following manner:

$$\begin{aligned} v \frac{\partial f}{\partial x} + \frac{d\varphi}{dx} \frac{df}{dv} &= 0, & \frac{d^2\varphi}{dx^2} &= \int_{-\infty}^{+\infty} f dv, \\ f(0, v) &= \Theta(v) \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2} + \varphi_1} \text{ for } v > 0, \\ \Theta(v) &= \begin{cases} 1 & \text{for } v \geq \sqrt{2V_1}, \\ 0 & \text{for } v < \sqrt{2V_1}, \end{cases} \\ f(l, v) &= 0 \quad \text{for } v < 0, \\ \varphi(0) &= \varphi_1, \quad \varphi(l) = \varphi_2, \end{aligned} \quad (6)$$

$l = d \sqrt{\frac{4\pi e^2 N}{kT}}$ is the dimensionless grid-reflector separation.

2. DIODE OPERATING CONDITION IN WHICH A VIRTUAL DIODE EXISTS (REGIME III)

In this case it is convenient to introduce a new system of coordinates, choosing the origin at the potential minimum. The potential will also be reckoned from the minimum. Denoting the new coordinates by ξ and the new potential by η , one may write

$$\xi = x - x_m, \quad \eta = \varphi - V_m, \quad (7)$$

where x_m is the position of the potential minimum relative to the grid; V_m is the depth of the minimum relative to the cathode. The approximate shape of the potential curve is illustrated in Figure 1.

In the new variables, system (6) takes the form

$$\begin{aligned} v \frac{\partial f}{\partial \xi} + \frac{d\eta}{d\xi} \frac{\partial f}{\partial v} &= 0, \quad \eta(\xi_{(1)}) = \eta_1, \quad \eta(\xi_{(2)}) = \eta_2, \\ \frac{d^2\eta}{d\xi^2} &= \int_{-\infty}^{+\infty} f dv, \quad f(\xi_{(1)}, v) = \Theta(v) \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2} + \varphi_1} \text{ for } v > 0, \\ f(\xi_{(2)}, v) &= 0 \quad \text{for } v < 0, \\ \Theta(v) &= \begin{cases} 1 & \text{for } v \geq \sqrt{2V_1}, \\ 0 & \text{for } v < \sqrt{2V_1}. \end{cases} \end{aligned} \quad (8)$$

Here $\xi_{(1)}$, $\xi_{(2)}$ and η_1 , η_2 are the coordinates and the grid and reflector potential respectively in the new system. It is obvious that

$$\begin{aligned} -\xi_{(1)} + \xi_{(2)} &= l, \\ \eta_1 - \eta_2 &= \varphi_1 - \varphi_2. \end{aligned} \quad (9)$$

Equation (9) together with the system of equations and boundary conditions (8) completely define the problem.

To solve system (8) we use methods developed by L. A. Vainshtein [5] and L. E. Pargamanik [6].

The general solution of the kinetic equation of system (8) is an arbitrary function of the magnitude $\frac{v^2}{2} - \eta(\xi)$:

$$f = f \left\{ \frac{v^2}{2} - \eta(\xi) \right\}. \quad (10)$$

A solution of Equation (10) must satisfy boundary conditions (8) and consequently, $f(\xi, \nu)$ should be different from zero only over specific velocity intervals.

In the diode space, in the case investigated, it is necessary to distinguish between three regions: $\eta_1 \leq \eta \leq \eta_1 - V_1$ first region; $\eta_1 - V_1 \leq \eta \leq 0$ second region; and $0 \leq \eta \leq \eta_2$ (Third region) (Figure 1).

For the second and third region, the solutions whose accuracy is of the order of the factor e^{φ_1} agree with the corresponding expressions of the theory of a diode with an incandescent cathode [5]. They can be obtained after integration of the system of Poisson equations (8).

For the second region

$$\frac{d\eta}{d\xi} = \left\{ -\frac{2e^{\varphi_1 - \eta_1}}{V\pi} \int_0^\infty e^{-\nu} (V\overline{y + \eta} - V\overline{y}) dy + \right. \\ \left. + \frac{4e^{\varphi_1 - \eta_1}}{V\pi} \int_{-\eta}^0 e^{-\nu} V\overline{y + \eta} dy \right\}^{+\frac{1}{2}}, \quad (11)$$

$$\xi'_1 = -e^{\frac{\eta_1 - \varphi_1}{2}} \int_0^\eta \left\{ \frac{2}{V\pi} \int_0^\infty e^{-\nu} (V\overline{y + \eta} - V\overline{y}) dy + \right. \\ \left. + \frac{4}{V\pi} \int_{-\eta}^0 e^{-\nu} V\overline{y + \eta} dy \right\}^{-\frac{1}{2}} d\eta. \quad (12)$$

For the third region

$$\xi_2 = e^{\frac{\eta_1 - \varphi_1}{2}} \int_0^\eta \left\{ \frac{2}{V\pi} \int_0^\infty e^{-\nu} (V\overline{y + \eta} - V\overline{y}) dy \right\}^{-\frac{1}{2}} d\eta. \quad (13)$$

Tables have been compiled [1] for values of ξ_1 and ξ_2 as a function of η .

We consider the solution for the first region. In this region electrons may have velocities within the following ranges:

$$\begin{aligned} \infty \geq v \geq \sqrt{2(V_1 - \eta_1 + \eta)} \quad \text{for } v > 0, \\ -\sqrt{2\eta} \geq v \geq -\sqrt{2(V_1 - \eta_1 + \eta)} \quad \text{for } v < 0. \end{aligned} \quad (14)$$

Consequently, for velocities within the ranges (14), the distribution function satisfying boundary condition (8) is

$$f(\xi, v) = \frac{1}{V^{2\pi}} e^{-\frac{v^2}{2} + \eta - \eta_1 + \varphi_1} \quad (15)$$

For all other velocities $f(\xi, v) = 0$.

The equation for the potential takes the form

$$\frac{d^2\eta}{d\xi^2} = \frac{e^{\varphi_1 - \eta_1}}{V^{2\pi}} \int_{\sqrt{2(V_1 - \eta_1 + \eta)}}^\infty e^{-\frac{v^2}{2} + \eta} dv + \frac{e^{\varphi_1 - \eta_1}}{V^{2\pi}} \int_{-\sqrt{2\eta}}^{-\sqrt{2(V_1 - \eta_1 + \eta)}} e^{-\frac{v^2}{2} + \eta} dv. \quad (16)$$

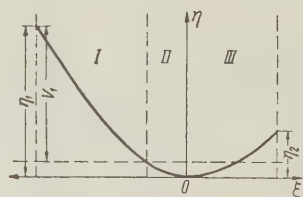


Figure 1. Qualitative picture of the diode potential distribution in the presence of a virtual cathode (regime III).

Making the change of variable $y = \frac{v^2}{2}$, we obtain

$$\frac{d^2\eta}{d\xi^2} = \frac{e^{\varphi_1 - \eta_1}}{\sqrt{2\pi}} \left\{ \int_0^\infty \frac{e^{-v} dy}{\sqrt{2(y+\eta)}} + 2 \int_{-(\eta_1 - V_1)}^0 \frac{e^{-v} dy}{\sqrt{2(y+\eta)}} \right\}. \quad (17)$$

Integrating Equation (17), we obtain

$$\begin{aligned} \frac{d\eta}{d\xi} = & \left\{ \frac{2e^{\varphi_1 - \eta_1}}{\sqrt{\pi}} \int_0^\infty e^{-v} (\sqrt{y+\eta} - \sqrt{y}) dy + \right. \\ & \left. + \frac{4e^{\varphi_1 - \eta_1}}{\sqrt{\pi}} \int_{-(\eta_1 - V_1)}^0 e^{-v} (\sqrt{y+\eta} - \sqrt{y}) dy + C \right\}^{\frac{1}{2}}. \end{aligned} \quad (18)$$

The integration constant C may be found from the continuity condition of the electric field at the boundary between the first and second region. Using Equations (11) and (18) we obtain

$$C = + \frac{4e^{\varphi_1 - \eta_1}}{\sqrt{\pi}} \int_{-(\eta_1 - V_1)}^0 e^{-v} \sqrt{y} dy. \quad (19)$$

Integrating Equation (18) and determining the constant from the potential continuity condition we finally obtain

$$\begin{aligned} \xi_1 = & -e^{\frac{\eta_1 - \varphi_1}{2}} \int_{\eta_1 - V_1}^\eta \left\{ \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-v} (\sqrt{y+\eta} - \sqrt{y}) dy + \right. \\ & \left. + \frac{4}{\sqrt{\pi}} \int_{-(\eta_1 - V_1)}^0 e^{-v} \sqrt{y+\eta} dy \right\}^{-\frac{1}{2}} d\eta - e^{\frac{\eta_1 - \varphi_1}{2}} \times \\ & \times \int_0^{\eta_1 - V_1} \left\{ \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-v} (\sqrt{y+\eta} - \sqrt{y}) dy + \frac{4}{\sqrt{\pi}} \int_{-\eta}^0 e^{-v} \sqrt{y+\eta} dy \right\}^{-\frac{1}{2}} d\eta. \end{aligned} \quad (20)$$

An equation determining the parameter η_1 may be obtained, by substituting values of ξ from Equations (20) and (13) into Equation (9). In developed form, this equation is

$$\begin{aligned} & e^{\frac{\eta_1 - \varphi_1}{2}} \int_{\eta_1 - V_1}^{\eta_1} \left\{ \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-v} (\sqrt{y+\eta} - \sqrt{y}) dy + \right. \\ & \left. + \frac{4}{\sqrt{\pi}} \int_{-(\eta_1 - V_1)}^0 e^{-v} \sqrt{y+\eta} dy \right\}^{-\frac{1}{2}} d\eta + \\ & - e^{\frac{\eta_1 - \varphi_1}{2}} \int_0^{\eta_1 - V_1} \left\{ \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-v} (\sqrt{y+\eta} - \sqrt{y}) dy + \right. \\ & \left. + \frac{4}{\sqrt{\pi}} \int_{-\eta}^0 e^{-v} \sqrt{y+\eta} dy \right\}^{-\frac{1}{2}} d\eta + e^{\frac{\eta_1 - \varphi_1}{2}} \times \\ & \times \int_0^{\eta_1 - (\varphi_1 - \varphi_2)} \left\{ \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-v} (\sqrt{y+\eta} - \sqrt{y}) dy \right\}^{-\frac{1}{2}} d\eta = l. \end{aligned} \quad (21)$$

These expressions determine the static characteristic of the diode for this case. The solution is applicable also for $\varphi_2 < 0$.

3. THE OPERATING CONDITION OF THE DIODE AT WHICH A POTENTIAL MINIMUM EXISTS, BUT A VIRTUAL CATHODE DOES NOT ARISE (REGIME II)

We find a solution for this case by the method used in the preceding paragraph. We distinguish between two regions of the diode (Figure 2): $0 \leq \eta \leq \eta_1$ (first region) and $0 \leq \eta \leq \eta_2$ (second region).

For the first region

$$f(\xi, v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2} + \eta_1 - \eta_1 + \varphi_1} \quad \text{for } \infty \geq v \geq \sqrt{2(V_1 - \eta_1 + \eta)},$$

$$f(\xi, v) = 0 \quad \text{for } v < \sqrt{2(V_1 - \eta_1 + \eta)}.$$
(22)

The equation for the potential agrees with Equation (16), if in Equation (16) we omit the second term of the right-hand side. After simple transformations analogous to those performed earlier, we obtain

$$\frac{d^2 \eta}{d\xi^2} = \frac{e^{\varphi_1 - \eta_1}}{\sqrt{2\pi}} \int_{V_1 - \eta_1}^{\infty} \frac{e^{-v} dy}{\sqrt{2(y + \eta)}}.$$
(23)

After integration of Equation (23) we find ξ as a function of η :

$$\xi_1 = -e^{\frac{\eta_1 - \varphi_1}{2}} \int_0^{\eta} \left\{ \frac{2}{\sqrt{\pi}} \int_{V_1 - \eta_1}^{\infty} e^{-v} (\sqrt{y + \eta} - \sqrt{y}) dy \right\}^{-\frac{1}{2}} d\eta.$$
(24)

A solution of the second region in no way differs from Equation (24) excluding $\xi_2 = -\xi_1$. It is not difficult to see that when $\eta_1 = V_1$ Equation (24) agrees with Equations (13) and (20).

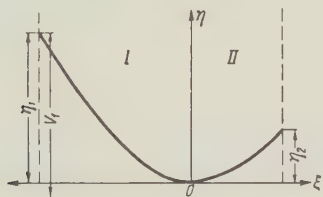


Figure 2. Qualitative picture of the potential distribution in a diode under the conditions that the potential minimum exists, but no virtual cathode emerges (regime II).

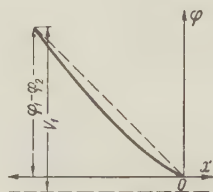


Figure 3. Qualitative picture of the potential distribution in the diode in the absence of a potential minimum. The case when $\varphi_2 \geq -\varphi_0$ (regime I).

The equation for determining η_1 which agrees with Equation (21) when $\eta_1 = V_1$ is

$$e^{\frac{\eta_1 - \varphi_1}{2}} \int_0^{\eta_1} \left\{ \frac{2}{\sqrt{\pi}} \int_{V_1 - \eta_1}^{\infty} e^{-v} (\sqrt{y + \eta} - \sqrt{y}) dy \right\}^{-\frac{1}{2}} d\eta + e^{\frac{\eta_1 - \varphi_1}{2}} \times$$

$$\times \int_0^{\eta_1 - (\varphi_1 - \varphi_2)} \left\{ \frac{2}{\sqrt{\pi}} \int_{V_1 - \eta_1}^{\infty} e^{-v} (\sqrt{y + \eta} - \sqrt{y}) dy \right\}^{-\frac{1}{2}} d\eta = l.$$
(25)

4. OPERATING CONDITION OF THE DIODE AT WHICH THE POTENTIAL DISTRIBUTION DOES NOT POSSESS A MINIMUM (REGIME I)

At first we consider the case when $\varphi_2 \geq -\varphi_0$. Let the origin of the coordinates coincide with the reflector plate (Figure 3). The potential φ also will be reckoned from the reflector. Then the solution of the kinetic equation satisfying the boundary conditions, assumes the following form:

$$f(x, v) = \frac{e^{\varphi_2}}{\sqrt{2\pi}} e^{-\frac{v^2}{2} + \varphi} \quad \text{for } \infty \geq v \geq \sqrt{2(V_1 - \varphi_1 + \varphi_2 + \varphi)},$$

$$f(x, v) = 0 \quad \text{for } v < \sqrt{2(V_1 - \varphi_1 + \varphi_2 + \varphi)}.$$
(26)

Substituting this function into Poisson's equation and making transformations analogous to the previous ones we obtain

$$\frac{d\varphi}{dx} = e^{\frac{\varphi_2}{2}} \left\{ \frac{2}{\sqrt{\pi}} \int_{V_1 - (\varphi_1 - \varphi_2)}^{\infty} e^{-y} \sqrt{y + \varphi} dy + C_1 \right\}^{\frac{1}{2}},$$
(27)

where C_1 is the integration constant. Integrating with respect to φ , we obtain x as a function of φ :

$$x = e^{-\frac{\varphi_2}{2}} \int_0^{\varphi} \left\{ \frac{2}{\sqrt{\pi}} \int_{V_1 - (\varphi_1 - \varphi_2)}^{\infty} e^{-y} \sqrt{y + \varphi} dy + C_1 \right\}^{-\frac{1}{2}} d\varphi.$$
(28)

The constant C_1 is determined from the condition

$$e^{-\frac{\varphi_2}{2}} \int_0^{\varphi_1 - \varphi_2} \left\{ \frac{2}{\sqrt{\pi}} \int_{V_1 - (\varphi_1 - \varphi_2)}^{\infty} e^{-y} \sqrt{y + \varphi} dy + C_1 \right\}^{-\frac{1}{2}} d\varphi = l.$$
(29)

At small electron densities in the beam $l \rightarrow 0$ and consequently, $C_1 \rightarrow \infty$. Equations (28) and (29) in this case lead to a linear potential drop:

$$\varphi = -\frac{\varphi_1 - \varphi_2}{l} x.$$
(30)

We note that Equation (29) must agree with Equation (25), if in Equation (25) we set $\eta_1 = \varphi_1 - \varphi_2$ (a potential minimum is found at the reflector; regimes I and II coincide). Consequently, in this case

$$C_1 = -\frac{2}{\sqrt{\pi}} \int_{V_1 - (\varphi_1 - \varphi_2)}^{\infty} e^{-y} \sqrt{y} dy.$$
(31)

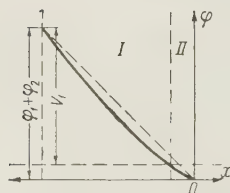


Figure 4. Qualitative picture of the potential distribution in the diode in the absence of the potential minimum. The case when $\varphi_2 < -\varphi_0$ (regime I).

$f(x, v) = 0$ for all the remaining values of v .

After substituting Equation (32) into the right-hand side of the equation for the potential and making the same transformations as in the preceding cases, we obtain

This value of C_1 is minimal for fixed V_1, φ_1 and φ_2 , because if the opposite were true l ceases to be real. If, in addition, $V_1 = \varphi_1 - \varphi_2$ (all three regimes coincide), $C_1 = -1$.

We now proceed to the case $\varphi_2 < -\varphi_0$. For this condition in the diode space, two regions should be distinguished (Figure 4): $\varphi_1 + \varphi_2 \geq \varphi \geq \varphi_1 + \varphi_2 - V_1$ (first region) and $\varphi_1 + \varphi_2 - V_1 \geq \varphi \geq 0$ (second region).

In the first region

$$f(x, v) = \frac{e^{-\varphi_2}}{\sqrt{2\pi}} e^{-\frac{v^2}{2} + \varphi}$$

$$\text{for } \infty \geq v \geq \sqrt{2(V_1 - \varphi_1 - \varphi_2 + \varphi)} \quad (32)$$

$$\text{and } -\sqrt{2(V_1 - \varphi_1 - \varphi_2 + \varphi)} \geq v \geq -\sqrt{2\varphi},$$

$$\frac{d\varphi}{dx} = e^{-\frac{\varphi_2}{2}} \left\{ \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-y} \sqrt{y+\varphi} dy + \frac{4}{\sqrt{\pi}} \int_{-(\varphi_1+\varphi_2-V_1)}^0 e^{-y} \sqrt{y+\varphi} dy + C_2 \right\}^{\frac{1}{2}}, \quad (33)$$

C_2 is the integration constant. Hence

$$x_1 = -e^{\frac{\varphi_2}{2}} \int_{\varphi_1+\varphi_2-V_1}^{\varphi} \left\{ \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-y} \sqrt{y+\varphi} dy + \frac{4}{\sqrt{\pi}} \int_{-(\varphi_1+\varphi_2-V_1)}^0 e^{-y} \sqrt{y+\varphi} dy + C_2 \right\}^{-\frac{1}{2}} d\varphi - x_{02}, \quad (34)$$

where x_{02} is the value of coordinate x at the point $\varphi = \varphi_1 + \varphi_2 - V_1$.

In the second region

$$f(x, v) = \frac{e^{-\frac{\varphi_2}{2}}}{\sqrt{2\pi}} e^{-\frac{v^2}{2} + \varphi} \quad \text{for } \infty \geq v \geq -\sqrt{2\varphi},$$

$$f(x, v) = 0 \quad \text{for } v < -\sqrt{2\varphi}. \quad (35)$$

The derivative of the potential corresponding to function (35) is

$$\frac{d\varphi}{dx} = e^{-\frac{\varphi_2}{2}} \left\{ \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-y} \sqrt{y+\varphi} dy + \frac{4}{\sqrt{\pi}} \int_{-\varphi}^0 e^{-y} \sqrt{y+\varphi} dy + C_2' \right\}^{\frac{1}{2}}. \quad (36)$$

It follows from the continuity of this derivative that $C_2' = C_2$. Integrating Equation (36) we obtain

$$x_2 = -e^{\frac{\varphi_2}{2}} \int_0^{\varphi} \left\{ \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-y} \sqrt{y+\varphi} dy + \frac{4}{\sqrt{\pi}} \int_{-\varphi}^0 e^{-y} \sqrt{y+\varphi} dy + C_2 \right\}^{-\frac{1}{2}} d\varphi. \quad (37)$$

Using Equations (34) and (37) we write an equation for determining C_2 :

$$e^{\frac{\varphi_2}{2}} \int_{\varphi_1+\varphi_2-V_1}^{\varphi_1+\varphi_2} \left\{ \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-y} \sqrt{y+\varphi} dy + \frac{4}{\sqrt{\pi}} \int_{-(\varphi_1+\varphi_2-V_1)}^0 e^{-y} \sqrt{y+\varphi} dy + C_2 \right\}^{-\frac{1}{2}} d\varphi + e^{\frac{\varphi_2}{2}} \int_0^{\varphi_1+\varphi_2-V_1} \left\{ \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-y} \sqrt{y+\varphi} dy + \frac{4}{\sqrt{\pi}} \int_{-\varphi}^0 e^{-y} \sqrt{y+\varphi} dy + C_2 \right\}^{-\frac{1}{2}} d\varphi = l. \quad (38)$$

Equation (38) should change to Equation (21) if $\eta_1 = \varphi_1 - \varphi_2$ and $\varphi_2 < 0$. This will happen when $C_2 = -1$. This value of C_2 is minimal.

5. CONDITIONS UNDER WHICH DIFFERENT DIODE OPERATING CONDITIONS EXIST

First of all it should be pointed out at what values of the retarding potential φ_2 different diode operating conditions can be realized.

A solution corresponding to a virtual cathode in the grid-reflector space, can hold at both positive and negative potentials φ_2 .

Regime II can be realized only when $\varphi_2 \geq -\varphi_0$, as may be seen directly from Figure 2. Regime I may exist both when $\varphi_2 > -\varphi_0$ and when $\varphi_2 < -\varphi_0$. The solutions for these cases have a different form.

We now proceed to the main question: at what initial current densities are the different

regimes realized, if the grid and reflector potentials are fixed? We consider the case of greatest interest $\varphi_2 > -\varphi_0$, when all three regimes may occur.

In the form of solution found by us it is convenient to consider no initial current density, and $l = d\sqrt{\frac{4\pi e^2 N_0}{kT}}$ is the dimensionless grid- to reflector separation, introduced previously.

In this expression N_0 is a magnitude directly proportional to the current passing through the grid:

$$j = eN_0 \sqrt{\frac{kT}{2\pi m}} e^{-\frac{e\varphi_0}{kT}}. \quad (39)$$

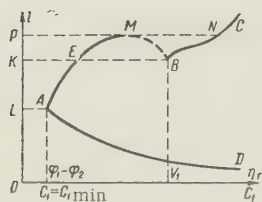


Figure 5. Approximate character of the dependence of the dimensionless grid-to-reflector separation on the parameters η_1 and C_1 for different diode operating conditions. Curve AD corresponds to regime I; AEMB-regime II; BC-regime III.

It is necessary to show at what values of the dimensionless distance l these regimes may be realized. The possible values of l at fixed φ_1 , φ_2 and V_1 are determined from Equations (21), (25), (29) and (38) which express the dependence of l on the parameters η_1 is apparent from Figures 1 and 2. For regime III $\eta_1 \geq V_1$; for regime II $\varphi_1 - \varphi_2 \leq \eta_1 \leq V_1$. In the case of regime I

$$C_1 \geq -\frac{2}{V\pi} \int_{V_1-(\varphi_1-\varphi_2)}^{\infty} e^{-y} V\bar{y} dy. \quad (40)$$

In Figure 5 curve BC illustrates the approximate variation of l with η_1 for regime III. According to Equation (21) l increases monotonically with η_1 . Consequently, regime III can be realized only when $l \geq OK$, i. e., under the condition

$$l \geq e^{\frac{\varphi_2}{2}} \int_0^{V_1} \left\{ \frac{2}{V\pi} \int_0^{\infty} e^{-y} (V\bar{y} + \eta - V\bar{y}) dy \right\}^{-\frac{1}{2}} d\eta + e^{\frac{\varphi_2}{2}} \int_0^{V_1-(\varphi_1-\varphi_2)} \left\{ \frac{2}{V\pi} \int_0^{\infty} e^{-y} (V\bar{y} + \eta - V\bar{y}) dy \right\}^{-\frac{1}{2}} d\eta. \quad (41)$$

Curve AB (Figure 5) illustrates the approximate behavior of $l(\eta_1)$ for regime II (according to the results of numerical integration). Of paramount importance is the existence of a maximum at point M. The possible values of l for regime II lie within the interval $OL \leq l \leq OP$. In Figure 5 the range of existence of regime I is indicated. According to Equation (29) l decreases monotonically as C_1 increases. Regimes I and II coincide when Equation (31) and the condition $\eta_1 = \varphi_1 - \varphi_2$ are fulfilled. Consequently, permissible values of l can be represented by curve AD. The first regime corresponds to values of l over the range $O \leq l \leq OL$.

Curves BC, AB and AD give the limits of existence of the different regimes and explain the hysteresis character of the current variation at the reflector as a function of initial current. At small currents regime I is realized. As the current increases continuously there occurs a transition to regime II without a current change at the reflector. At current values corresponding to $l = OP$ (point M on curve AB), the diode jumps in steps to regime III (point N on curve BC) and the reflector current drops abruptly because a virtual cathode is formed. With decreasing initial current, the diode operates in regime III down to $l = OB$, when the diode changes by stages to regime II (point E on curve AB) a transition is also possible in regime I at several values of φ_1 and φ_2 . In this case there will be no abrupt jump in the reflector current, because at $\eta_1 = V_1$ the current of regime II equals the current of regime III. Section MB of the curve cannot be realized.

Such a hysteresis type character of the reflector current should be observed also in

the case of a constant current through the grid, and in the case of a changing potential φ_2 .

We note that the experimentally observable current jumps have a sharply pronounced character only during a diode transition from regime II into regime III. During an inverse transition the current changes rather quickly, but in spite of everything smoothly. This result agrees qualitatively with the theory expounded above. We note again that when $\varphi_2 < -\varphi_0$ regime II is not realized and, consequently there should be no hysteresis relation for the current. This conclusion is also found to be in accord with experiment.

In conclusion I thank A.I. Kostienko, R. L. Stratonovich and M. N. Devyatkov for their interest in the work and useful discussions.

REFERENCES

1. S. D. Gvozdozer, Theory of Electron Devices at Very-High Frequencies, pp. 96-134, GTTI, 1956.
2. A. I. Kostienko, M. N. Devyatkov, A. A. Lebed', Radiotekhnika i Elektronika, 1959, 4, 3, 482.
3. G. Ya. Myakishev, Vestn. Mosk. Univyerzityeta, 1960, 3, 5, 19.
4. A. A. Vlasov, Theory of Many Particles, GTTI, 1950.
5. L. A. Vainshtein, Theory of shot effect in the presence of space charge, Sb. Nauchnykh Trudov, vyp. XI, Izd. Sovyetskoye Radio, 1948.
6. L. E. Pargamanik, ZhETF, 1957, 33, 1, 251.

M. B. Lomonosov Moscow State University
Physics Faculty

Submitted to the editors 15 June 1960

TRANSITION PROCESSES IN SEMICONDUCTOR DIODES USING NARROW FORWARD CURRENT PULSES

Yu. R. Nosov

We consider pulse parameters describing the transition process, when switching a diode from the forward to the reverse direction. Relations are obtained for these parameters as a function of the duration of forward current flow.

INTRODUCTION

In papers devoted to a consideration of the process of switching a semiconductor diode from a conducting to a nonconducting state, the duration of forward current flow is usually assumed so large that the hole distribution in the diode base * may be regarded as in the steady-state [1, 2, 3].

In practice, however the use of diodes in fast-acting circuits requires us to frequently operate with extremely narrow forward current pulses. Consequently, it is of interest to calculate how one may vary the parameters describing diode transition characteristics in the case of a nonsteady state hole distribution in the base.

On the other hand a development of the technique of measuring pulse parameters of semiconducting diodes often requires us to transmit through the diode extremely narrow

*Hereinafter it is assumed that the diode base is an n-type conductor, so that its minority carriers are holes.

forward current pulses before switching (in pulse generators an increase of the steepness of the rise time is connected, as a rule, with a reduction of the possible pulse length). For this purpose it is important to determine the minimum duration of the forward current necessary to achieve a steady-state distribution of holes in the diode base.

The goal of this paper also is to clarify the influence of the duration of a forward current pulse on the character of the transient process upon switching the diode from the forward to the reverse direction.

1. FUNDAMENTAL ASSUMPTIONS AND EQUATIONS

We shall consider the model of a one-dimensional junction diode with an abrupt p-n junction and with a semi-infinite base region, the p-region conductivity being assumed much larger than the n-region conductivity (see Figure 1). Just as in References 1, 2, 3, the injection level is assumed to be small, but the lifetime of holes in the base is assumed to be a constant. The influence of the electrostatic capacitance of a p-n junction on the character of transient processes is excluded.**

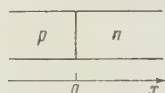


Figure 1. Model of a one-dimensional junction diode.

These assumptions are satisfied, to some extent, by domestic industrial types of pulse diodes which are pseudopoint structures: (for example, diodes D9, D11-14, D101-103). Microjunction silicon and germanium quick-acting diodes conform well with this model.

The behavior of holes in the diode base is described by the diffusion equation

$$\frac{\partial p}{\partial t} = D_p \frac{\partial^2 p}{\partial x^2} - \frac{p - p_{n0}}{\tau_p}, \quad (1)$$

where p is the density; D_p is diffusion coefficient; τ_p is the lifetime; p_{n0} is the equilibrium density of holes in the base.

Neglecting the magnitude of p_{n0} in comparison with p and introducing dimensionless time $\Theta = t/\tau_p$ and distance $X = x/L_p$ variables (L_p is the hole diffusion length in the base) we obtain a diffusion equation in the form

$$\frac{\partial p}{\partial \Theta} = \frac{\partial^2 p}{\partial X^2} - p. \quad (1a)$$

The current density through a p-n junction is expressed by the following equation:

$$j = - \frac{qD_p}{L_p} \frac{\partial p}{\partial X} \Big|_{X=0} \quad (2)$$

In the steady state, the hole distribution in the base has the form

$$p = p_{n0} e^{qU_f/kT} e^{-x/L_p} = p_0 e^{-\lambda x}, \quad (3)$$

Where U_f is the forward bias on the diode; q is the electron charge; k is Boltzmann's constant; T is the absolute temperature; p_0 is the hole concentration near the p-n junction in the steady state. Consequently, for a steady-state current flow we have

$$j = \frac{qD_p p_0}{L_p}. \quad (4)$$

The solution of the problem formulated in the introduction is broken down into two stages: 1) a calculation of the establishment of a hole distribution in the base, upon connecting the diode in the forward direction; 2) a calculation of the junction diode transient characteristic upon switching it to the reverse direction taking into account the distribution obtained.

**Nosov [4] has more thoroughly formulated the basic hypotheses underlying the modern theory of transition processes in semiconductor diodes.

2. DETERMINATION OF THE HOLE DISTRIBUTION DURING THE PASSAGE OF A FORWARD CURRENT PULSE

A step in the forward current density from zero to j_f leads to the emergence of a hole concentration gradient close to p-n junction given by

$$\left. \frac{\partial p}{\partial X} \right|_{X=0} = j_f \frac{I_n}{q D_p} = p_0, \quad t = 0. \quad (5)$$

Up to the starting moment of a forward pulse, the hole concentration in the diode base is negligibly small:

$$p(X, 0) = 0, \quad t = 0. \quad (6)$$

We solve the diffusion equation (1a), assuming the initial distribution (6) and boundary condition (5). With the substitution

$$p = e^{-\Theta U}(X, \Theta) \quad (7)$$

Equation (1a) reduces to the form

$$\frac{\partial U}{\partial \Theta} = \frac{\partial^2 U}{\partial X^2}. \quad (8)$$

Integrating Equation (8) by means of a two-dimensional Laplace-Carson transformation [5] and taking into account the substitution (7), we obtain

$$p = p_0 \frac{1}{2} \left\{ e^{-X} \operatorname{erfc} \left(\frac{X}{2\sqrt{\Theta}} + \sqrt{\Theta} \right) + e^X \operatorname{erfc} \left(\frac{X}{2\sqrt{\Theta}} - \sqrt{\Theta} \right) \right\}^*. \quad (9)$$

Figure 2 gives graphs of the relation $\frac{p}{p_0}(X)$ for different fixed values of Θ . From solution (9) it is easy to obtain

$$p_{X=0} = p_0 \operatorname{erf} \sqrt{\Theta}. \quad (10)$$

Hence the voltage rise at a p-n junction is given by

$$U_f(\Theta) = \frac{kT}{q} \ln \frac{p_{X=0}}{p_{n0}} = U_{f-s} - \frac{kT}{q} \ln [\operatorname{erf} \sqrt{\Theta}]^{-1},$$

where U_{f-s} is the voltage at the p-n junction in the case of a steady-state hole distribution in the base.

Because U_{f-s} is usually of the order of 0.25-0.5 v and at room temperature $kT/q = 0.025$ v, the voltage at the p-n junction reaches 0.8 of its steady-state value, under the condition $\ln [\operatorname{erf} \sqrt{\Theta}] = -(2-4)$, which leads to $\Theta = 0.02-0.004$. So small a value of Θ explains the fact that no steady-state voltage at the p-n junction is experimentally observed, because it usually takes place during a time less than the resolving time of the apparatus.

An examination of the graphs of Figure 2 shows that when $\Theta \ll 1$ the hole distribution in the base may be described by an exponential relation of the form

$$p(X, \Theta) = p(\Theta) e^{-X l(\Theta)}. \quad (12)$$

We will define $p(\Theta)$ and $l(\Theta)$. Using a decomposition into a series of error integrals, we obtain from Equation (10)

$$p(\Theta) = p_0 \frac{2}{\sqrt{\pi}} \sqrt{\Theta}. \quad (13)$$

*Here and hereinafter it is assumed that $z = t - \operatorname{erf} z \equiv t - \frac{2}{\sqrt{\pi}} \int_0^t e^{-v^2} dv$.

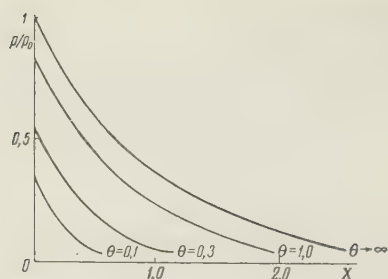


Figure 2. Distribution of the hole density in the diode base for various durations of a forward current pulse.

Integrating Equation (1) over the volume of the base and noting that $q \int_0^\infty p dx = Q_H$ is the total hole charge accumulated in the base per unit area of the p-n junction, we obtain the following equation for this process of accumulation:

$$Q_H(\Theta) = Q_H(\Theta \rightarrow \infty)(1 - e^{-\Theta}), \quad (14)$$

where $Q_H(\Theta \rightarrow \infty) = qp_0L_p$. When $\Theta \ll 1$ we have

$$Q_H(\Theta) = Q_H(\Theta \rightarrow \infty)\Theta = qp_0L_p\Theta. \quad (15)$$

Integration of Equation (12) multiplied by q over the base region gives

$$Q_H(\Theta) = p(\Theta)l(\Theta)q. \quad (16)$$

Equating expressions (15) and (16) and taking into account Equation (13), we obtain

$$l(\Theta) = \frac{V\pi}{2} \frac{1}{V\Theta} L_p. \quad (17)$$

Thus the hole distribution in the diode base at $\Theta \ll 1$ can be described by the expression

$$p(X, \Theta) = p_0 \frac{2}{V\pi} V\Theta e^{-\frac{2}{V\pi} \frac{1}{V\Theta} X}. \quad (18)$$

Numerical calculations carried out at $\Theta = 0.1$ show that Equations (18) and (9) differ by no more than 3% for arbitrary X values.

Letting $2/\sqrt{\pi} \simeq 1$ as an approximation we note that the hole distribution in the diode base for a small forward pulse duration ($t \ll \tau_p$) is the same as for the steady-state case if the hole concentration p_0 near the p-n junction and the diffusion length L_p are decreased by a factor of $1/\sqrt{\Theta}$.

At the end of this paragraph, we will find the establishment of the hole distribution in the diode base during the instantaneous transmission of forward voltage* at the p-n junction**, i.e., we will integrate Equation (1a) with condition (6) and the boundary condition

$$p|_{X=0} = p_0 \text{ for } t > 0. \quad (19)$$

The solution is obtained just as for the case of constant current and is

$$p(X, \Theta) = \frac{p_0}{2} \left\{ e^{-X} \operatorname{erfc} \left(\frac{X}{2V\Theta} - V\Theta \right) + e^X \operatorname{erfc} \left(\frac{X}{2V\Theta} + V\Theta \right) \right\}. \quad (20)$$

The current density through the p-n junction during the process of establishment is

$$j = j_f \left[\frac{e^{-\Theta}}{(\pi\Theta)^{1/2}} - \operatorname{erf} V\Theta \right] \quad (21)$$

and at small values of Θ equals

$$j = j_f \frac{1}{V\pi\Theta}. \quad (21a)$$

*The magnitude of the forward voltage is taken so that, during establishment of equilibrium, the same current density j_{f-s} flows through the diode as in the previous discussion.

**Such a calculation was made in Reference 6 for a diode with a thin base.

The accumulation of hole charge in the diode base is described in this case by the expression

$$\rho_n(\Theta) = \int_0^t j dt = \frac{2}{V\pi} Q_H(\Theta \rightarrow \infty) V\Theta. \quad (22)$$

Comparison of Equations (15) and (22) shows that the rate of accumulation of holes in the diode base, immediately after switch-on (i.e., $\Theta \ll 1$), upon applying a voltage pulse to the p-n junction, is $1/\sqrt{\Theta}$ times larger than for the case of a current pulse. Physically this result is explained by the fact that the pulse strength of the current density during the process of attaining equilibrium, at all times, is larger than the steady-state current density.

It should be noted that in real electron circuits the diode forward resistance, as a rule, is much less than the load resistance connected in series with it, i.e., the conditions for a current generator are realized. Therefore, in future considerations we use the solution obtained with boundary condition (5), but not (19).

Graphs of $\frac{p(X)}{p_0}$ for different fixed values of Θ , given in Figures 2 and 3, show that with small Θ the distributions $p(X, \Theta)$ for the case of a switch-over to the voltage generator regime are substantially different.

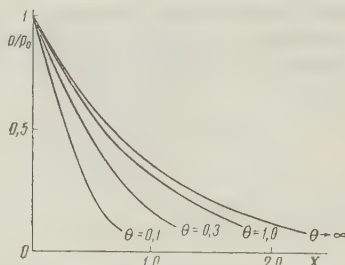


Figure 3. Distribution of hole density in the diode base for different forward voltage pulse durations.

3. SWITCHING OF A DIODE FROM FORWARD TO REVERSE DIRECTION

When switching a diode from the forward to the reverse direction, a reverse voltage pulse is applied to the p-n junction and the series resistance R (equal to the total resistance of the diode base and load).

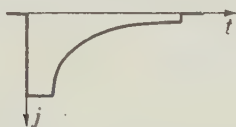


Figure 4. Theoretical transition characteristic of a diode when switching it from the forward to the reverse direction.



Figure 5. Transition characteristic of the diode at a finite rate of switching.

The form of the transition characteristic for this process is shown in Figure 4. The limiting case for $R \rightarrow 0$ [3] leads to the disappearance of the "shelves" and to an infinitely large value of reverse current at the instant of switching. Two circumstances, however, substantially change the initial stage of the switching process.

The finiteness of the maximum possible velocity v_{\max} limits the hole current through the p-n junction [7]. This hole current is directly proportional to the hole concentration near the p-n junction $p_{x=0}$. The other limitation on reverse current is related to the finite rate of any real switching. In this case the maximum value of reverse current is also determined by the magnitude of $p_{x=0}$, although the direct proportionality is violated.

Because of these limitations, a real transition characteristic of the diode has the form shown in Figure 5.

A quantitative calculation of the pulsed properties of a diode is made via the introduction of three parameters:

1. The magnitude of the maximum value of reverse current during switching I_{\max} which characterizes the initial stage of the transition process.

2. Recovery time of reverse diode resistance τ_{rec} determined by the time interval elapsing from the beginning of switching up to the instant the diode reverse current drops to a certain prescribed value. The recovery time of the diode back resistance usually equals (3-4) τ_p and thus characterizes the "tail" part of the transition process.

3. The magnitude of electric charge passing through the diode in the reverse direction during switching Q_B .

If we assume that $I_{\text{max}} \sim p_{x=0}$ is correct, then

$$I_{\text{max}}(\Theta_0) = I_{\text{max}}(\Theta_0 \rightarrow \infty) \operatorname{erf} \sqrt{\Theta_0}; \quad I_{\text{max}}(\Theta_0) = 0,8 I_{\text{max}}(\Theta_0 \rightarrow \infty) \\ \text{for } \Theta_0 = 0,8.$$

In order to determine the dependence of τ_{rec} on the duration of a forward current pulse Θ_0^* it is necessary to find the entire diode transfer characteristic.

We will integrate diffusion equation (1) taking the initial distribution

$$p(X, 0) = p(X, \Theta_0) = \frac{p_0}{2} \left\{ e^{-X} \operatorname{erfc} \left(\frac{X}{2\sqrt{\Theta_0}} - \sqrt{\Theta_0} \right) - e^X \operatorname{erfc} \left(\frac{X}{2\sqrt{\Theta_0}} + \sqrt{\Theta_0} \right) \right\} \quad (9a)$$

and the boundary condition

$$p(0, \Theta) = 0, \quad \Theta > 0. \quad (23)$$

Assumption of the boundary condition (23) indicates that a back voltage is instantaneously applied to the p-n junction. This assumption is correct, if it is considered removed from the initial stage of the transition process (as in our case) and the resistance in the diode circuit is so small that the length of the "shelf" determined by this resistance is insignificant**. In most cases using pulse diodes, the assumption of condition (23) is most reasonable.

From the general solution of the problem, obtained using the original function for a semi-infinite region, we find

$$\frac{\partial p}{\partial X} \Big|_{X=0} = \frac{e^{-\Theta}}{2\pi^{1/2}\Theta^{1/2}} \int_0^\infty p(X, \Theta_0) e^{-X^2/4\Theta} X dX. \quad (24)$$

We transform the initial distribution (9a) to the form

$$p(X, \Theta_0) = p_0 e^{-X} - \frac{p_0}{2} \left\{ e^{-X} \operatorname{erfc} \left(\sqrt{\Theta_0} - \frac{X}{2\sqrt{\Theta_0}} \right) + \right. \\ \left. + e^X \operatorname{erfc} \left(\sqrt{\Theta_0} + \frac{X}{2\sqrt{\Theta_0}} \right) \right\} = p_0 e^{-X} - \varphi(X, \Theta_0). \quad (9b)$$

Solution (22) at $p(X, 0) = \varphi(X, \Theta_0)$ is discussed in detail in Reference 3, where it is shown that for large values of Θ we have

$$\frac{e^{-\Theta}}{2\pi^{1/2}\Theta^{1/2}} \int_0^\infty \varphi(X, \Theta_0) e^{-\frac{X^2}{4\Theta}} X dX = \frac{p_0 e^{-\Theta}}{2\pi^{1/2}\Theta^{1/2}} \left[\frac{2\Theta_0^{1/2} e^{-\Theta_0}}{\pi^{1/2}} + \operatorname{erfc} \Theta_0^{1/2} \right]. \quad (25)$$

Under this same condition

$$\frac{e^{-\Theta}}{2\pi^{1/2}\Theta^{1/2}} \int_0^\infty p_0 e^{-X} e^{-\frac{X^2}{4\Theta}} X dX = p_0 \frac{e^{-\Theta_0}}{2\pi^{1/2}\Theta_0^{1/2}}. \quad (26)$$

*Hereinafter the duration of the forward current pulse Θ_0 is a parameter, but Θ is a variable magnitude measured from the beginning of the switching.

**The subject of the quantitative relation between the length of the "shelf" and the diode circuit resistance is considered in Reference 4.

Taking into account Equations (24), (25) and (26) the density of the diode reverse current at large Θ^* is expressed by

$$j_{rev} = j_f \frac{1}{p_0} \left(\frac{\partial p}{\partial X} \right)_{X=0} = j_f \frac{e^{-\Theta}}{2\pi^{1/2}\Theta^{1/2}} \left[1 - \left(\frac{2\Theta^{1/2}e^{-\Theta_0}}{\pi^{1/2}} + \operatorname{erfc} \Theta^{1/2} \right) \right]. \quad (27)$$

For a large duration of forward current ($\Theta_0 \rightarrow \infty$) this relation has the form [3]

$$j_{rev} = j_f \frac{e^{-\Theta}}{2\pi^{1/2}\Theta^{1/2}}. \quad (28)$$

The graphs of Fig. 6 illustrate Equation (27). It is easy to see that the recovery time of the back resistance reaches 0.8 of its steady-state value in the case of a forward current pulse length of $\Theta_0 \simeq 0.8$.

Thus the transition process of switching both in the initial (specific parameter I_{max}) stages and in the "tail" stages (determined by the parameter τ_{rec}) are practically independent of the forward current duration, if it exceeds $0.8 \tau_p$.

To determine the amount of charge passing through the diode in the reverse direction during switching, one must know the entire transition characteristic which for arbitrary values of Θ_0 was not found by us.

We find the transition characteristic of a diode for extremely narrow forward current pulses ($\Theta_0 \ll 1$), using the approximate initial hole distribution (18)**.

The solution found in this manner will differ from that for the case of the actual initial distribution, however, by virtue of the correctness of the statement of the boundary problem of diffusion equation [8], the difference between these solutions will be no larger than the difference between solutions (18) and (19). Carrying out an incomplete integration of the diffusion equation, we find by means of an operational method [5], the value of the gradient of the hole concentration near the p-n junction

$$\left. \frac{\partial p}{\partial X} \right|_{X=0} = \frac{2p_0 \sqrt{\Theta_0}}{\sqrt{\pi}} \left[\frac{e^{-\Theta}}{\sqrt{\pi\Theta}} - \frac{2}{\sqrt{\pi\Theta_0}} e^{-\Theta \left(1 - \frac{4}{\pi\Theta_0} \right)} \operatorname{erfc} \frac{2\sqrt{\Theta}}{\sqrt{\pi\Theta_0}} \right] \quad (29)$$

Using Equations (2) and (29), we determine the amount of charge passing through the diode in the reverse direction after switching,

$$Q_H(\Theta_0) = \int_0^\infty j_{rev} dt = qp_0 L_p \frac{\Theta_0}{1 + \frac{\Theta_0}{2}} = Q_H(\Theta_0 \rightarrow \infty) \frac{\Theta_0}{1 + \frac{\Theta_0}{2} \sqrt{\pi\Theta_0}}, \quad (30)$$

where $Q_H(\Theta_0 \rightarrow \infty)$ is the total charge of the holes accumulated in the diode base. In view of Equation (15), we obtain

$$\xi(\Theta_0) = \frac{Q_B(\Theta_0)}{Q_H(\Theta_0)} = \frac{1}{1 + \frac{\Theta_0}{2} \sqrt{\pi\Theta_0}} \simeq 1 - \frac{\sqrt{\pi}}{2} \sqrt{\Theta_0}. \quad (31)$$

This calculated expression for $\xi(\Theta_0)$ determines the correction to the pulse method of measuring the injection coefficient γ [9] and should be considered when setting up appropriate experiments.

*Calculations show that when $\Theta > 3$ the error is less than 20% in comparison with $\Theta \rightarrow \infty$.

**Here, as previously, boundary condition (23) is assumed.

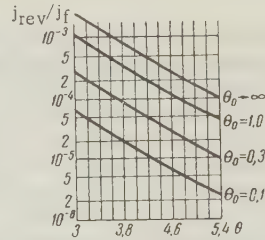


Figure 6. Variation of j_{rev}/j_f with Θ .

As is well-known, when $\Theta_0 \rightarrow \infty$ $\xi(\Theta_0) \rightarrow 1/2$ and, consequently, the possible values of $\xi(\Theta_0)$ lie between 1 and 1/2. Since the total charge Q_H passing through the diode during the passage of the forward current pulse equals $Q_p = I_f t_0 = Q_H(\Theta_0 \rightarrow \infty) \Theta_0$, taking into consideration Equation (14), we obtain

$$\eta = \frac{Q_b}{Q_p} = \xi \frac{1 - e^{-\Theta_0}}{\Theta_0}. \quad (32)$$

The coefficient η^* characterizes the rectifying ability of the diode when using it in the switching regime. As is apparent from Equation (32), when $\Theta_0 = 1$ $\eta(\Theta_0) = 0.3 - 0.5$ and only when $\Theta_0 = 5$ $\eta(\Theta_0) = 0.1$. This means that the rectifying ability of the diode during operation with narrow pulses of forward current ($\Theta_0 \leq 1$) gets considerably worse.

CONCLUSIONS

1. Pulse parameters which characterize switching of the diode from the forward to the reverse direction, achieve virtually their steady-state values at a forward current pulse duration of $t_f > 0.8 \tau_p$.
 2. During a forward current pulse, the establishment of the voltage at the p-n junction occurs during the time $t \leq 0.02 \tau_p$.
 3. With narrow pulses of forward current ($t_f \ll \tau_p$) the hole distribution in the base has an exponential character, wherein the coefficient of the exponential term and the exponent are decreased by a factor of $\sqrt{\pi}/2\sqrt{\Theta_0}$ in comparison with a steady state distribution.
 4. When measuring γ it is necessary to introduce the coefficient $\xi = 1 - \frac{\sqrt{\pi}}{2} \sqrt{\Theta_0}$, taking into account the hole recombination in the diode base.
 5. The rectifying ability of the diode which operates in the switching regime, gets abruptly worse at a duration of the forward current pulse $t_f \leq \tau_p$.
- The author expresses his gratitude to K.S. Rzhevkin and B.G. Mendelev for a discussion of the work and a number of valuable comments.

REFERENCES

1. R.H. Kingston, Proc. I.R.E., 1954, 42, 5, 829.
2. B. Lax, S.F. Neustadter, J. Appl. Phys., 1954, 25, 9, 1148.
3. I.G. Henderson, I.P. Tillman, Proc. I.E.E., 1957, 104, p. B, 15, 318.
4. Yu.R. Nosov, Transition Characteristics of Semiconductor Diodes. Coll. Semiconductor Devices and their Application, Editor Ya.A. Fedotova, Izvest. Sovetskoye Radio 4, p. 3, 1959.
5. V.A. Ditkin and A.P. Prudnikov, Operational Calculus with Respect to Two Variables and its Application, Fizmatgiz, 1958.
6. L.M. Baranov and M.S. Bekbulatov, Radiotekhnika i Elektronika, 1959, 4, 4, 703.
7. T.R. Scott, Proc. I.E.E., 1957, 107, p. B, 15, 333.
8. S.L. Sobolev, Equations of Mathematical Physics, GITTL, 1954.
9. O.L. Curtic, B.R. Gossick, Rev. Scient. Instrum., 1956, 27, 10, 828.

Received by the Editors 23 May 1960

*This magnitude is analogous to the diode rectification factor (the ratio of forward to reverse current), which characterizes the diode rectifying properties in the static regime.

MOLECULAR FREQUENCY STANDARD WITH SUBTRACTION OF REFERENCE OSCILLATOR ERROR

V. V. Grigor'yants and M. Ye. Zhabotinskiy

The report presents a brief discussion of a number of circuits providing standard frequencies on the basis of a molecular oscillator. Calculations are performed for a frequency standard with subtraction of the error of a reference oscillator which provides stable frequencies in the centimeter, decimeter and meter wave ranges. Circuits are developed for investigation of frequency-phase stability of the obtained standard signal: (1) by means of an auxiliary molecular oscillator; (2) by means of an auxiliary quartz oscillator and frequency multiplier.

The advent of the molecular oscillator [1, 2], opening wide prospects for the creation of highly stable frequency and time standards, called for the development of methods of transferring its stability into the range of radio frequencies and at the same time increasing its power. Various authors have recently proposed a large number of circuits permitting practical utilization of the high stability of the molecular oscillator. Disregarding purely technical differences, these circuits may be divided into three main groups according to their principle of operation.

I. Circuits for frequency comparison of the molecular oscillator and an auxiliary oscillator (usually using a quartz crystal) permitting highly accurate measurement of the departure of the latter from the desired frequency [3, 4, 9].

II. Circuits for automatic control of the auxiliary oscillator (by means of the molecular oscillator), differing from the circuits of group I in the presence of a control loop which acts upon the auxiliary oscillator (automatic frequency or phase control) [5, 6, 7, 8].

III. A circuit for the creation of a standard frequency by subtraction of the instability of a reference oscillator (see Figures 1 and 2), proposed by the authors in 1955 [9]. This scheme possesses a number of advantages over those in the above groups: first, in the absence of feedback, which eliminates the problem of system stability as a whole; second, in the absence of influence on the quartz oscillator, which ensures a purer standard frequency spectrum. However, in its initial form this circuit had a number of significant disadvantages.

1. Due to the considerable difficulty in obtaining local-oscillator power directly from a frequency multiplier, (see Figure 1) it is necessary to use an auxiliary local oscillator (see Figure 2) having sufficiently high frequency stability, determined by the passband of the i-f amplifier (100 kc). This leads to complication of the circuit and to additional phase modulation of the effective signal.

2. For excellent filtering the subtracting sections must possess high equivalent Q.

3. The circuit provides only one stable frequency.

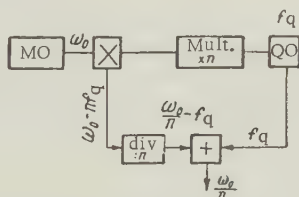


Figure 1

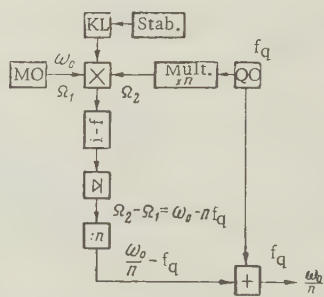


Figure 2

In conjunction with these disadvantages it was undertaken to develop an improved version of the circuit with error subtraction. The problem was approached in three phases: first, to eliminate the auxiliary local oscillator; second, to improve filtering and the subtraction sections; third, to obtain frequencies in the centimeter, decimeter and meter wave ranges

simultaneously. The work resulted in the design of a frequency standard with subtraction of the reference oscillator error which meets the above requirements (see Figure 3).

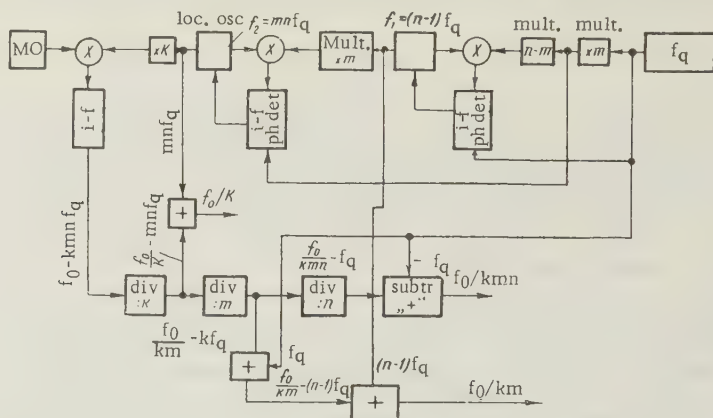


Figure 3

As is seen from the diagram, the problem of obtaining local-oscillator power is solved by automatic phase control of the klystron f_1 through a quartz oscillator f_q and automatic phase control of klystron f_2 through klystron f_1 with use of the same oscillator f_q as the reference oscillator. Use of the phase stabilization circuit mentioned permits subtraction of the error at three frequencies simultaneously: f_2 , f_1 and f_q . In addition, construction and alignment of the automatic phase control circuit is considerably simplified by virtue of the considerable signal levels from the mixers (the mixer at frequency f_1 provides a signal of several tens of millivolts, the mixer at frequency f_2 provides a signal of several hundred millivolts). This automatic phase control circuit is equivalent to a high-power frequency multiplier, since phase drifts of the automatic phase control are small [10].

Let us examine the relationships which must be fulfilled in this circuit.

For operation of the automatic phase control of f_1 through f_q with the use of f_q as the reference signal it is necessary that

$$f_1 = (n \pm 1) f_q. \quad (1)$$

It is shown below that in Equation (1) it is best to choose the minus sign:

$$f_1 = (n-1)f_q. \quad (2)$$

Further, for automatic phase control of f_2 through f_1 let us choose the following frequency relationship:

$$f_2 = mf_1 + mf_q \quad (3)$$

or, using Equation (2),

$$f_2 = (mn - m + m)f_q = mnf_q. \quad (4)$$

Frequency f_2 is chosen in the region of the k -th subharmonic of the molecular oscillator, that is,

$$f_{i-f} = f_0 - kmnf_q. \quad (5)$$

Let us now discuss subtraction of the error at frequencies f_2 , f_1 and f_0 .

Subtraction at frequency f_2 . In this case the intermediate frequency of Equation (5) must be divided by the factor k . Then adding f_2 , we obtain

$$\frac{f_0 - kmnf_q}{k} + f_2 = \frac{f_0}{k} - mnf_q + mnf_q = \frac{f_0}{k}. \quad (6)$$

In this case, a stable frequency is obtained in the centimeter wave range, for in practice, due to the necessity of obtaining from the f_2 oscillator a local-oscillator power in the range of the molecular oscillator, $k = 1, 2, 3$.

Subtraction at frequency f_1 . The division factor of f_{1-f} must be equal to km . Then adding f_1 , we obtain

$$\frac{f_0 - kmnf_q}{km} + f_1 = \frac{f_0}{km} - nf_q + (n - 1)f_q = \frac{f_0}{km} - f_q \quad (7)$$

Hence, it is evident that in order to obtain a stable frequency it is necessary to add f_q :

$$\frac{f_0}{km} - f_q + f_q = \frac{f_0}{km} \quad (8)$$

With $m = 3$ to 10 we obtain a stable frequency in the decimeter range.

Subtraction at frequency f_q . The division factor of f_{1-f} must be equal to kmn . Then adding f_q , we obtain

$$\frac{f_0 - kmnf_q}{kmn} + f_q = \frac{f_0}{kmn} \quad (9)$$

With appropriate choice of n we obtain a stable frequency in the meter range. It is now evident why the minus sign is chosen in Equation (1) and in Equation (3) $f_2 - mf_q$. It is this choice of frequency ratios which has ensured a signal in the automatic phase control circuit equal to $kmn f_q$, which in turn permits eliminating from the circuit of the standard, one frequency divider with successive division factors of k , m and n . Moreover, an additional advantage resulting from Equations (2) and (3) is the difference in the intermediate frequencies and, correspondingly, in the reference signals in the automatic phase control of f_1 through f_q and in the circuit for automatic phase control of f_2 through f_1 , which decreases the influence of parasitic coupling in the circuit. For tube economy and the compensation of phase drifts of the frequency multiplier of the quartz oscillator it is desirable that the multiplier have a multiplication factor equal to m . Then from the point of the multiplier where the signal of frequency mf_q exists a tap may be made to the phase detector of the automatic phase control circuit of f_2 through f_1 ; otherwise, frequency mf_q must be specially created.

We shall now discuss the requirements for the subtraction sections. For the sake of simplicity in evaluation we shall consider sideband filtering at the -3db level to be adequate. To the subtracting device (assuming this is an ordinary mixer) there is applied the signal from the divider with a frequency of f_{1-f}/m_1 , where m_1 is the division factor, and a reference oscillator signal with frequency kq . The distance between the side frequencies is $2f_{1-f}/m_1$, that is, the Q of the filter for attenuation of the side frequencies at the -3db level is

$$Q = \frac{m_1 f_q}{2f_{1-f}} \quad (10)$$

Hence it is evident that in order to improve the filtering it is necessary to decrease the division factor, preceding subtraction, and to increase f_{1-f} .

Equation (10) may be rewritten in another form if we consider that $m_1 f_q = f_0 - f_{1-f}$ and $f_{1-f} \ll f_0$, where f_0 is the frequency of the molecular oscillator,

$$Q = \frac{f_0 - f_{1-f}}{2f_{1-f}} \simeq \frac{f_0}{2f_{1-f}} \quad (11)$$

Consequently, in ordinary subtraction the required filter Q does not depend on the waveband at which subtraction is performed. Let us define this Q Equation (11), assuming $f_{1-f} = 50$ Mc:

$$Q = \frac{2.4 \cdot 10^4}{10^2} = 240,$$

that is, in the centimeter range where cavity resonators with Q of $(3 \text{ to } 5) \cdot 10^4$ are available, we may obtain side-frequency attenuations of 10^2 to 10^3 times. The Q of cavities in the decimeter range is lower, hence in filtering with a single tuned circuit we may obtain side-frequency attenuations of only 10 to 100 times. In the meter range it is difficult to ensure the required filtering with a single tuned circuit, but this difficulty is overcome by the use of the simplest automatic phase control circuit. In addition, early in 1958 S. P. Dubonosov proposed that subtraction of the reference oscillator error be performed in two successive steps, that is, actually to lower the value of m_1 in Equation (10). This method permitted satisfactory filtering of side frequencies in any waveband. Above, in the examination of error subtraction at frequency f_1 dual subtraction was performed automatically, otherwise it would not be possible to obtain stable frequency f_0/k . In this manner filtering in this range was also facilitated (see Equations (8) and (7)).

It should be noted that dual subtraction of error is sometimes unsatisfactory. In such cases it is necessary to change to higher order subtraction. In this connection let us examine sequential subtraction of reference oscillator error of different orders up to the general case of the k -th order subtraction.

Let us first discuss the relation between the frequency multiplication factor l ($l = kmn$) in Equation (5) of the quartz reference oscillator (which yields the frequency of the molecular oscillator) and the division factors m_k for various orders of error subtraction. In this case let us consider that $f_{1-f} = f_0 - lf_q$.

(1) Single subtraction

$$\frac{f_0 - lf_q}{m_1} + f_q = \frac{f_0}{m_1},$$

if $l = m_1$.

(2) Double subtraction:

$$\frac{f_0 - lf_q}{m_1 m_2} + \frac{f_q}{m_2} + f_q = \frac{f_0}{m_2 m_1},$$

if $l = m_1(1 + m_2)$.

(3) Triple subtraction:

$$\frac{f_0 - lf_q}{m_1 m_2 m_3} + \frac{f_q}{m_2 m_3} + \frac{f_q}{m_3} + f_q = \frac{f_0}{m_1 m_2 m_3},$$

if $l = m_1(1 + m_2 + m_2 m_3)$.

(4) k -th subtraction:

$$\frac{f_0 - lf_q}{m_1 m_2 \dots m_k} + \frac{f_q}{m_k m_{k-1} \dots m_2} + \dots + \frac{f_q}{m_k} + f_q = \frac{f_0}{m_1 m_2 \dots m_k},$$

if $l = m_1(1 + m_2 + m_2 m_3 \dots m_k)$,

that is, in the general case

$$l = m_1(1 + m_2 + m_2 m_3 + \dots + m_2 m_3 \dots m_k) \quad (12)$$

or with $m \simeq 10$

$$l \simeq m_1 m_2 \dots m_k. \quad (12a)$$

Let us now discuss the requirements imposed on the Q of the filters of the subtraction sections according to the subtraction order. As before, for the sake of simplicity let us assume that side-frequency division at the -3db level is adequate.

(1) Single subtraction.

Let us rewrite Equation (10), for $f_{1-f} = pf_q$ where p is some coefficient

$$Q_1 = \frac{m_1 f_q}{2f_{1-f}} = \frac{m_1 f_q}{2pf_q} = \frac{m_1}{2p} \simeq \frac{m_1}{2}.$$

(2) Double subtraction:

$$Q_2 = \frac{m_2 f_q}{2 \left(\frac{f_{i-f}}{m_1} + f_q \right)} = \frac{m_1 m_2 f_q}{2 f_q (p + m_1)} \approx \frac{m_2}{2},$$

(3) Triple subtraction:

$$Q_3 = \frac{m_3 f_{KB}}{2 \left(\frac{f_{i-f}}{m_1 m_2} + \frac{f_q}{m_2} + f_q \right)} = \frac{m_1 m_2 m_3 f_q}{2 f_q (p + m_1 + m_1 m_2)} \approx \frac{m_3}{2}.$$

(4) k-th subtraction:

$$Q_k = \frac{m_k f_q}{2 \left(\frac{f_{i-f}}{m_1 m_2 \dots m_k} + \frac{f_q}{m_2 \dots m_k} + \dots + f_q \right)} = \frac{m_1 m_2 \dots m_k f_q}{2 (p + m_1 + \dots + m_1 m_2 \dots m_{k-1}) f_q} \approx \frac{m_k}{2}. \quad (13)$$

In the above simplifications it was assumed that $p \approx 1$ and all $m \approx 10$. If needed, the precise equation (13) may be used (up to the \approx sign), particularly in those cases where f_{i-f} and f_q are widely different in value (that is, $p \gg 1$ or $p \ll 1$).

In the specific calculations of any scheme for a frequency standard based on the principle of subtraction of the reference oscillator error it is necessary to assign two initial values: (1) the level of side frequencies in the output signal, which permits determining Q_k and m_k from Equation (13); (2) the frequency of the output signal, knowing which, we may approximately determine the value of l .

Further, knowing l and m_k and assuming in Equation (12a) that $m_1 = m_2 = \dots = m_k$, let us determine the required error subtraction order from the formula

$$l \approx m^k. \quad (12b)$$

After this it is necessary to refine the value of l and all m_k from Equation (12) and the value of f_q from the formula $f_{i-f} = f_0 - l f_q$.

METHOD FOR MEASURING THE STABILITY OF FREQUENCY AND PHASE OF THE STANDARD SIGNAL

The usual method of studying frequency standards is by comparison of several identical circuits (the so-called internal comparison).

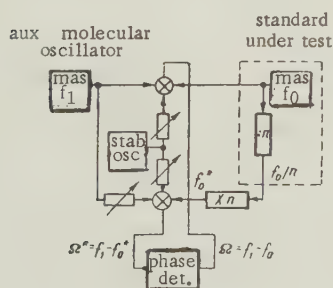


Figure 4

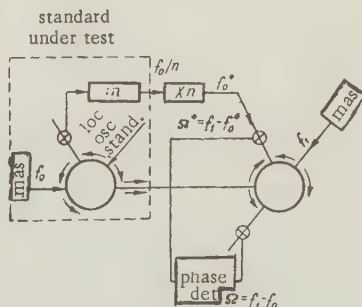


Figure 5

Below we examine a circuit for comparison of the output signal with the auxiliary molecular oscillator.

The circuit develops a standard signal with frequency $f_0/n = f_c$. In studying this signal it is most simple, multiplying it n times, to compare its n -th harmonic $f_{cn} = f_0^*$ with the signal of a second molecular oscillator f_1 . In this case we obtain the difference frequency

$$\Omega^* = f_1 - f_0.$$

Included in Ω^* , as also in f_0^* , are the phase errors of the frequency standard circuit as well as the phase errors introduced by the auxiliary frequency multiplier.

For measurement of the frequency stability of the output signal of the investigated standard it suffices to compare Ω^* with the frequency of an auxiliary low-frequency oscillator. In order to evaluate stability with an error on the order of 10^{-10} the auxiliary oscillator may have an error on the order of $10^{-2} - 10^{-3}$.

In order to evaluate phase fluctuations in f_0^* it is, moreover, necessary to shift the signals of both molecular oscillators. In this case we obtain $\Omega = f_1 - f_0$.

Applying Ω^* and Ω to the phase detector, we may measure the phase fluctuations of the frequency-standard circuit.

The block diagram of a system based on the above considerations is shown in Figure 4. The stability of the auxiliary local oscillator must be such that it is desirable to employ quartz-controlled automatic phase control. Otherwise, extra phase modulation of Ω^* and Ω is possible.

As seen from Figure 3, the standard-frequency circuit includes a local-oscillator klystron whose phase is quartz stabilized. If the power of this local oscillator at wavelength of 1.25 cm is sufficiently great, it may be used at the same time in the circuit for measuring the stability of the frequency standard. The problem is further simplified if we do not require balanced reception of the molecular oscillator signal. The resulting loss in receiver sensitivity is quite small.

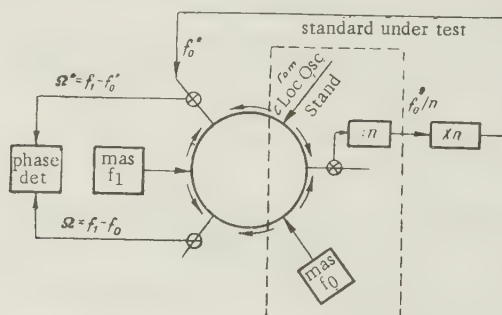


Figure 6

Two versions of the circuit using a local-oscillator frequency standard and unbalanced mixers are shown in Figures 5 and 6. The circuit in Figure 5 uses two hybrid rings of ordinary design. The circuit in Figure 6 uses one hybrid ring with six inputs instead of the usual four.

For measurement of the phase stability of the standard signal, in addition to the circuit described above, there was developed a circuit which generally does not require a molecular oscillator. A block diagram of this circuit is shown in Figure 7. The frequency of a quartz-controlled auxiliary oscillator (equal to the frequency f_0/n of the output signal of the frequency standard with error subtraction) is multiplied by a factor n and applied to its input as a simulator of the molecular oscillator. Accuracy of the output frequency of the standard is equal to that of the auxiliary quartz oscillator. By applying these two frequencies to the phase detector we may measure the phase stability of the standard signal. The requirements for stability of the quartz oscillator ($\sim 10^{-7}$) are determined by the i-f bandwidth, taking into account the frequency multiplication factor of the quartz crystal. The molecular oscillator simulator is also quite convenient for

prealignment of the standard-frequency circuit.

The circuit described above (Figure 7) was used for investigation of an experimental model of a frequency standard with subtraction which provided a stable frequency in the meter wave range. As was to be expected, the converted signal of the standard maintained the frequency stability of the molecular oscillator.

Random phase modulation of the signal (measured by means of the circuit shown in Figure 7) proved to be on the order of one angular minute.

The circuit of the frequency standard was tested in a setup with an independent molecular oscillator with a carbon absorption pump. The oscillator operated without interruption for more than 40 hours.

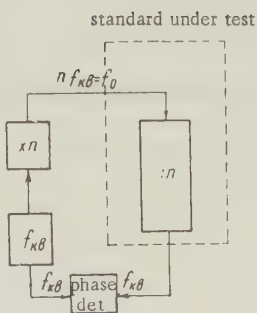


Figure 7

REFERENCES

1. N.G. Basov, *Pribery i tekhnika eksperimenta*, 1957, 1, 1-7.
2. T.P. Gordon, H.Z. Leiger, S.N. Townes, *Phys. Rev.*, 1954, 95, 282.
3. A. Ya. Leykin, Report at the Scientific Session of the A.S. Popov Society, May 1957, *Izmeritel'naya tekhnika*, 1959, 7, 41.
4. N.G. Basov, I.D. Murin, A.P. Petrov, A.M. Prokhorov, I.V. Shtranikh, *Izv. vuzov MVO SSSR (Radiofizika)*, 1958, 2, 3, 50.
5. I.D. Murin, *Izv. vuzov MVO SSSR (Radiotekhnika)*, 1957, 1, 5, 555.
6. G.A. Vasneva, B.A. Gaygerov, V.V. Grigor'yants, G.A. Yelkin, M. Ye., *Zhabotinskiy, Radiotekhnika i elektronika*, 1957, 2, 10, 1300.
7. E.T. Davis, Phase stabilization to microwave frequency standards, *Wescon reports*, August 1957.
8. I.L. Bershteyn, *Dokl. na konferentsii po statisticheskoy fizike*, Gorkiy, October 1958.
9. G.A. Vasneva, V.V. Grigor'yants, M. Ye. Zhabotinskiy, *Clock-maser circuits*, *Dokl. na 2-y Vsesoyuznoy konferentsii MVO SSSR po radioelektronike*, Saratov, September 1957.
10. I.L. Bershteyn, Report at the Scientific Session of the A.S. Popov Society, May 1957. *Radiotekhnika i elektronika*, 1958, 3, 2, 288; I.L. Bershteyn, V.L. Sibiryakov, *Radiotekhnika i elektronika*, 1958, 3, 2, 290.

Institute of Radio Engineering and
Electronics AN SSSR

Submitted to the editors 1 March 1960

BRIEF COMMUNICATIONS

WIDEBAND AMPLIFICATION WITH FREQUENCY CONVERSION

L. K. Mikhaylovskiy

The author has experimentally observed the interaction of two microwave signals in a nonlinear element, which interaction occurs at any arbitrary relationship between the signal frequencies.* This permitted proposing the creation of wideband microwave amplifiers based on a new method of amplification [1]. This method is based on control of the amplified signal by the parameters of the nonlinear element,** which parameters, in turn, cause a change in value of the complex reflection coefficient from the given element for the auxiliary (supply) signal. In addition, as has been shown by theory and experiment, it is possible to achieve amplification of the microwave envelope upon conversion of the carrier frequency.***

The essentials of the proposed method are as follows. Let two microwave signals be applied simultaneously to a given nonlinear element. One of the signals is a control signal whose envelope is to be amplified and the other is a continuous (or pulse) signal from an auxiliary generator.

Let the first signal have a carrier frequency f_c and mean (pulse) power P_c and let the second signal, correspondingly, have frequency f_g and power P_g . The signal to be amplified, in acting on the nonlinear element (e.g., a crystal detector in a detector head or a magnetized ferrite within a waveguide), causes a corresponding change in the capacitance and resistance of the crystal detector and the occurrence of precession of the magnetic moments of electron spins in the ferrite. Thus, the control signal is first, as it were, detected by the substance of the nonlinear element and then, as a consequence, causes a change in the external electromagnetic parameters of this nonlinear element for the auxiliary signal. The change in internal microparameters occurs in accordance with the change in envelope of the control signal but with a definite relaxation time which, in combination with the Q of the internal circuit of the nonlinear element, determines the amount of distortion of the envelope after its "transfer" to a new carrier.

The auxiliary signal, of course, also causes corresponding changes in the parameters of the nonlinear element and hence, during the action of the first signal, it should possess constant power ($P_{g\text{ appl.}} = \text{const}$ or $P_{g\text{ pulse appl.}} = \text{const}$). Hence it follows that the operating conditions of the nonlinear element must be so chosen that with a change in value of the auxiliary signal power reflected from the nonlinear element the parameters of this element must remain practically constant.

As a result of the above process it develops that, after the arrival of the control signal at the nonlinear element, amplitude and phase modulation corresponding to the envelope of the amplified signal appear in the reflected part of the auxiliary signal or that part of the auxiliary signal which has passed through the element.

*This effect, detected at microwavelengths in an artificial medium (a solid body), probably has much in common with the Luxemburg-Gor'kov effect observed at longer wavelengths in a natural medium — the ionosphere.

**By a nonlinear element is meant any device or substance changing its electromagnetic parameters under the influence of electromagnetic waves.

***Here and henceforth we refer to power amplification of the signal, for voltage amplification (or current amplification) has no significance in microwave engineering.

Mathematically this process may be described as follows. With $P_C = 0$, let the auxiliary signal power reflected from the nonlinear element be $\rho_2 (f_g P_g) P_g$ and with $P_C \neq 0$ let it be $\rho_2 (f_C f_g P_C P_g) P_g$, where ρ_1 and ρ_2 are the corresponding moduli of the complex reflection coefficients for the auxiliary signal, depending both on frequencies f_C and f_g and on the powers P_C and P_g [$\rho = (\frac{SWR - 1}{SWR + 1})^2$]. Moreover, it is evident that the difference

$$\rho_1 P_g - \rho_2 P_g = \Delta \rho P_g = \Delta P_g \quad (1)$$

will determine the value of the change in power of the auxiliary signal reflected from the nonlinear element due to the action of the amplified signal on this element. The gain in this case will be

$$k = \frac{\Delta P_g (f_C f_g P_C P_g)}{P_C} = \Delta \rho \frac{P_g}{P_C} \quad (2)$$

If the nonlinear element permits the application of sufficient power P_g that even at extremely small values of $\Delta \rho$ it develops that $\Delta \rho P_g > P_C$, there will then occur amplification of the signal envelope of f_C with conversion of its carrier to the frequency f_g .

It is evident that upon the action of a signal with power P_g (with $P_C = 0$) on the nonlinear element the signal reflected from this element will contain components with frequencies of $k f_g$, where $k = 1, 2, 3$, etc. If, in addition, this element is acted upon by a signal with power P_C , then the signal reflected from it will contain combination frequencies upon which there will be expended a fraction of power ΔP_g (with $\rho_2 < \rho_1$). The remaining part of ΔP_g will go into the increase in losses in the nonlinear element and the possible creation of oscillations of those frequencies for which the element, under the action of signal P_C , will become a negative resistance, the modulus of which is equal to or greater than the characteristic impedance of the waveguide. In the case where the power of these oscillations exceeds the value of P_C there also arises the effect of amplification of the signal envelope with conversion of its carrier frequency.

With $\rho_2 > \rho_1$ the power increment ΔP_g as well as the combination frequencies and the frequencies from the negative-resistance effect may obviously arise only due to a decrease in the power losses of the auxiliary signal in the nonlinear element and due to a decrease in the power of its harmonics $k f_g$ (except power P_C).

The described method of amplification is based on the change in value of the modulus of the corresponding reflection coefficient. It is evident that, in principle, equal success may be obtained by using the change in phase of the reflection coefficient and the change (under the influence of the control signal) of the corresponding transfer constant. In the latter case the nonlinear element will be a four-terminal network [2].

For an increase in k , nonlinear elements similar to those described above may be connected in parallel or in series; for an increase in the signal-to-noise ratio the auxiliary generator must be pulse triggered and automatically follow the signal to be amplified.

The frequency ranges of signals interacting by the method described above are determined by the properties of the nonlinear element and, in principle, may lie in any portion of the electromagnetic spectrum.

By way of example let us point out that the electrically and magnetically controlled microwave attenuators and phase inverters which have recently come into wide use are based on the effect of a change in the parameters of the substance of the nonlinear element under influence of the envelope of a control signal, with the frequency of this signal $f_C = 0$ and the frequency of the controlled signal f_g lying in the centimeter range.* If we examine the other extreme case with $f_g = 0$, when, instead of the microwave supply signal, there is applied to the nonlinear element a supply signal in the form of a d-c voltage and, possibly, a permanent constant magnetic field, then we likewise obtain amplification if under these conditions and with appropriate tuning the negative input resistance of the nonlinear element is ensured. In other words, a reflection coefficient greater than unity will be ensured for the amplified

*The possibility of creating electrically controlled (rather than mechanically controlled) oscillators was reported by the author in 1953-1954 at a seminar at Moscow State University conducted by S.D. Gvozdozer.

signal. Moreover, amplification will occur without change in the carrier frequency. Also, the closer the absolute value of the negative input resistance to the value of the characteristic impedance of the waveguide, the greater the gain. In the limiting case (that is, in the case of equality of these two values) oscillation will occur.

Cross modulation of the two microwave signals by the use of a nonlinear element was achieved under experimental conditions wherein the frequencies of both signals lay in the centimeter* and millimeter ranges and also with the signal of frequency f_c in the millimeter range and the signal of frequency f_g in the centimeter range.

In all of the above cases it was possible to distinguish cross modulation with amplification observed on an oscillograph screen. An experiment was also performed for the purpose of measuring k with both interacting signals in the centimeter range. The measurements were made with a thermistor power-measuring device. The value of k in using one nonlinear element under the specified conditions was approximately 40.

I wish to take this opportunity to express my thanks to B.P. Pollak and V.F. Balakov for their assistance in the experiments at millimeter wavelengths and to A.V. Zaytsev for assistance in the last of the above experiments.

REFERENCES

1. L.K. Mikhaylovskiy, Author's claim of 3 May 1957, No. 572528, before the Committee on Inventions and Discoveries, Council of Ministers USSR; author's certificate No. 114707, class 21g 13/17.
2. L.K. Mikhaylovskiy, Cross modulation in ferrites, Treatise III of the Conference on Ferrite Physics, Minsk, June 1959 (see the book Ferrites, Izd. AN BSSR, Minsk, 1960).

Submitted to the editors July 12, 1960

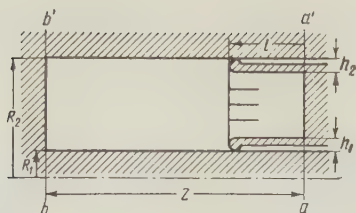
SUPPRESSION OF HARMONICS IN COAXIAL RESONATORS

A.I. Yel'kind

For resonators and filters with square-law tuning in the microwave range coaxial butterfly circuits are used which have a near-linear dependence of resonant wavelength λ_1 on the variable length Z of the resonator. A resonator of ideal shape, formed by the intersection of a coaxial cavity by two planes, has an absolutely linear characteristic but is usually not used due to the prevalence of harmonics. Upon tuning such a resonator to the fundamental frequency of the excitation signal, harmonic resonance may occur, which is inadmissible in most applications. In practical resonator designs the elimination of multiple frequencies and, consequently, of harmonic resonance is associated with one or another deviation from the ideal shape. Such deviation results in nonlinearity of the relationship $\lambda_1 = \lambda_1(Z)$ and from this point of view must be a minimal deviation. Hence it is of interest to determine the minimum geometric dimensions of the deviation eliminating harmonic resonance in a resonator tuned to the fundamental frequency of the signal.

Since the natural functions and frequencies for a coaxial resonator of ideal shape are known, such a problem is solved most simply by the perturbation method [1]. We present below an example of the calculations for a widely used type of resonator (see figure) with a

*In addition, a completely similar effect of cross modulation was established experimentally, even as a result of the second harmonic of the reflected auxiliary signal.



contacting shorting plunger. The deviation is introduced by the shape of the shorting plunger. The dimension h is usually small in comparison with R , which permits considering the resonator at significant values of l to be slightly deformed relative to the ideal resonator delimited by planes aa' and bb' . The spectrum of the natural frequencies of a resonator of length Z

$$\omega_n = \omega_{n0} + \delta\omega_n, \quad (1)$$

where

$$\omega_{n0} = n\omega_{10} = n\pi c / Z \quad (2)$$

is the spectrum of the undeformed resonator;

$$\delta\omega_n = \omega_{n0} \frac{\int_V (|H_n|^2 - |E_n|^2) dv}{2 \int_V |H_n|^2 dv} = \frac{\omega_{n0}}{2n\pi} f(R, h) \sin \frac{2n\pi}{Z} l; \quad (3)$$

$$H_n = H_0 \frac{R_1}{r} \cos n\pi z / Z; \quad E_n = iE_0 \frac{R_1}{r} \sin n\pi z / Z$$

are the eigenfunctions of the undeformed resonator ($E_0 = H_0$). The numerator is integrated over the deformation volume: for φ from 0 to 2π ; for z from 0 to l ; for r from R_1 to $R_1 + h_1$ and from $R_2 - h_2$ to R_2 — wherein, due to the smallness of h , we use $E(r) \approx E_0$ and $E(r) = E_0 R_1 / R_2$. Integrating the denominator over the entire volume:

$$f(R, h) = \left(\frac{h_1}{R_1} + \frac{h_2}{R_2} \right) \ln^{-1} \frac{R_2}{R_1}; \quad \ln \frac{R_2}{R_1} = \frac{W}{60},$$

where W is the characteristic impedance of the coaxial resonator in ohms.

In resonator operation the fundamental frequency of the signal coincides with the first natural frequency ω_1 . For the frequency of the n -th harmonic of the signal $n\omega_1$ detuning occurs relative to the nearest natural frequency ω_n . From Equations (1)–(3) it follows that

$$n\omega_1 - \omega_n = n\delta\omega_1 - \delta\omega_n = (\omega_{n0} / 2\pi) f(R, h) \left(\sin 2\pi l / Z - \frac{1}{n} \sin 2n\pi l / Z \right). \quad (4)$$

The equation $n \sin x - \sin nx = 0$, where $x = 2\pi l / Z$, has no roots in the interval $0 < x < \pi$. This indicates that for all harmonics the detuning differs from zero at resonator lengths $Z > 2l$. With $Z = 2l$ harmonic resonance is possible, but this case is remote from the practical application of resonators. For complete elimination (suppression) of harmonics the detuning must extend beyond the half-bandwidth of the resonance curve $\omega_n / 2Q_n$. In a resonator tuned to the fundamental frequency of the signal, harmonic resonance can occur only at the higher harmonics (n half-wavelengths of z). In this case the Q for the harmonics is higher than for the fundamental frequency and increases with n . $Q_n \approx \sqrt{n} Q_1$ [1].

It follows from Equation (4) that detuning also increases with the harmonic number. Consequently, it suffices to determine the condition for suppression of the lowest (second) harmonic. The minimum detuning ensuring practically complete suppression of the second harmonic is represented by the value

$$2\omega_1 - \omega_2 = \omega_2 / Q_1. \quad (5)$$

Such detuning exceeds by 2–3 times the half-bandwidth of the resonant curve $\omega_2 / 2Q_2$ ($Q_2 \approx 1.4 Q_1$)

The dimensions ι , h which yield detuning in accordance with Equation (5) or greater must satisfy the condition

$$\left(\frac{h_1}{R_1} + \frac{h_2}{R_2}\right)\left(\frac{\iota}{Z}\right)^3 \geq \frac{W}{120\pi^2 Q_1}, \quad (6)$$

which is obtained by substituting Equation (4) into Equation (5) and expanding Equation (4) into a series in terms of ι/Z up to the third order; ι and h satisfying Equation (6) are the minimum dimensions ensuring suppression of the second harmonic and, consequently, of any of the higher harmonics. With $W = 50$ ohms, $Q_1 = 10^3$ and $h/R = 0.1$ condition (6) is fulfilled with $\iota/Z > 0.04$. With smaller ι/Z or h/R suppression of harmonics is not ensured. Since the shape of the shorting plunger is usually not chosen for the purpose of the suppression of harmonics, the other end of the resonator (the bb' end) is given a specially graduated shape with the ι and h dimensions calculated from (6).

Taking into account the nonlinearity of $\lambda_1 = \lambda_1(Z)$, the tuning relationships for the resonator are obtained from Equations (1), (2) and (3), where $n=1$, $\lambda_1 = 2\pi c/\omega_1$.

REFERENCES

1. A.G. Gurevich, Cavity Resonators and Waveguides, Izd. Sovetskoye radio, 1952, pp. 94-97, 107.

Submitted to the editors April 28, 1960

CALCULATION OF RIBBON STREAM ELECTRON GUNS

L. E. Bakhrakh and A. D. Sulimin

1. SHAPE OF THE ELECTRODES

We will proceed from that part of the cylindrical diode where we use as a cathode the part of the internal side of the external cylinder bounded by a sector of angle 2θ (Figure 1). In a cylindrical diode with a space-charge limited current, the behavior of the potential $U(r)$ can be represented in the form

$$U(r) = \left[\frac{\frac{r}{r_K} \beta^2\left(\frac{r}{r_K}\right)}{\frac{r_a}{r_K} \beta_a^2\left(\frac{r_a}{r_K}\right)} \right]^{1/3} U_a, \quad (1)$$

where r_a and r_K are the anode and cathode radii respectively; r is the distance from the origin of the polar coordinates to the point under consideration in the interelectrode space; U_a is the anode potential; β is a function of the ratio r/r_K determined by the expression

$$\beta = \gamma - \frac{2}{5} \gamma^2 + \frac{11}{120} \gamma^3 - \frac{47}{3300} \gamma^4 + \dots;$$

$$\gamma = \ln \frac{r}{r_K}.$$

In order to preserve the rectilinear motion of electrons in this sector (under the condition that inside this sector Poisson's equation holds, but outside it, Laplace's equation holds), one must place outside the sector, electrodes of such a form that at the boundary of

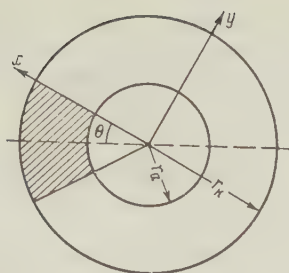


Figure 1

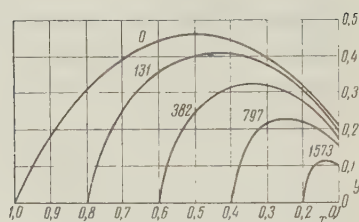


Figure 2

of the beam, the following condition is fulfilled:

$$U_{y=0} = U_0(x), \quad \left(\frac{\partial U}{\partial y} \right)_{y=0} = 0 \quad (2)$$

(i.e., at $y = 0$ in the x, y coordinate system (Figure 1), where $U_0(x)$ is the potential distribution corresponding to Equation (1).

We seek a solution of Laplace's equation in series form

$$U = \sum_{n=0}^{\infty} U_n(x) y^n. \quad (3)$$

On the basis of boundary conditions (2) we have

$$U(x, 0) = U_0(x), \quad \left(\frac{\partial U}{\partial y} \right)_{y=0} = U_1 = 0. \quad (4)$$

To find the other coefficients of the series we substitute Equation (3) into Laplace's equation, whence we obtain the recursion formula

$$U_{n+2}(x) = - \frac{U_n''(x)}{(n+2)(n+1)},$$

by means of which together with Equation (4) all the coefficients of the series can be expressed in terms of the function $U_0(x)$ and its derivatives.

The coefficients calculated in this manner allow a determination of $U(x, y)$ and consequently, the sought-for form of the electrodes. The result of similar calculations limited to three terms of the expansion is given in Figure 2. The electrode shape found in this

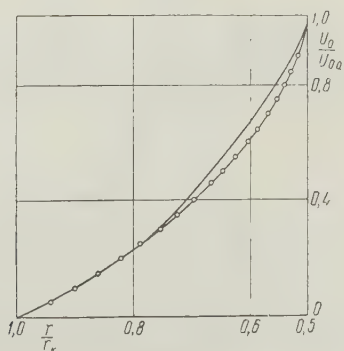


Figure 3

manner differs somewhat both from the electrode shape determined by Pierce [1] and from the calculation of Radley [2]. Figure 3 gives the behavior measured in an electrolytic tank of the potential along the boundary of the flow in the case of the electrode shape calculated above. One may see that the potential distribution obtained agrees satisfactorily with the theoretical (solid curve) corresponding to Equation (1).

2. EFFECT OF THE ANODE APERTURE AND SPACE CHARGE ON THE FORM OF THE FLUX IN THE ANODE SPACE

It is well-known that the space-charge limited current in a cylindrical diode of length is determined by expression

$$I'_a = \frac{14,64 \cdot 10^{-6} U_a^{3/2}}{r_a^2 \beta_a^2 \left(\frac{r_a}{r_k} \right)},$$

where $I'_a = I_a / l$ is the current per unit length along the axis of the cylinder in amperes per meter; r_a is the anode radius in meters; U_a is the anode voltage in volts. Consequently the perveance of such a cylindrical gun with an angle of convergence 2Θ (Θ in degrees) will be

$$P = \frac{I'_a}{U_a^{3/2}} = \frac{14,64 \cdot 10^{-6}}{r_a^2 \beta_a^2} \left(\frac{2\Theta}{360} \right). \quad (5)$$

A diaphragm with an aperture may be considered as a dispersing electron lens with a focal length determined from the relation

$$f = - \frac{2U_a}{E_a},$$

where E_a is the electric field intensity near the anode determined by Equation (1).

Using the lens equation

$$\frac{1}{r_a} - \frac{1}{s_0} = \frac{1}{f}, \quad (6)$$

where S_0 the distance from the lens plane to the point at which continuations of the electron trajectories would converge in the absence of the dispersing force of the space charge, one may reach the conclusion [4] that the lens action of the anode aperture is converging only for the values $r_a/r_k < 0.53$.

To explain the behavior of the electron stream in the anode space under the action of space charge, we will use the equation of motion of the outer electron. For a constricted stream half-width of y_m and for a constricted stream-to-anode aperture separation of z_m this equation gives [3]

$$y_m = y_0 \left(1 - \frac{y_0}{4Ks_0^2} \right), \quad (7)$$

$$z_m = \frac{y_0}{2Ks_0}, \quad (8)$$

where $K = \frac{I'_a}{U_a^{3/2}} = 2.4 \cdot 10^{-4}$; y_0 is the half-width of the stream in the anode aperture. At $K \leq y_0/4s_0^2$ the electron stream intersects the central plane, i.e., no constriction is formed and, consequently it is impossible to obtain a parallel electron stream at the output of the gun. Thus,

$$K \geq \frac{y_0}{4s_0^2} \quad (9)$$

is the condition for the presence of a constriction in the electron stream issuing from the gun. It is easy to see that this relation interrelates the quantities r_a , θ , r_a/r_k and P .

The solid curves of Figure 4 illustrate the variation of Pr_a with r_a/r_k at different values of the parameter θ according to Equation (5). The dashed curves show the relation between these same quantities according to Equation (9), in which we set $y_0 = r_a \tan \theta$. The intersection of these curves at the same values of θ gives a system of points lying on the

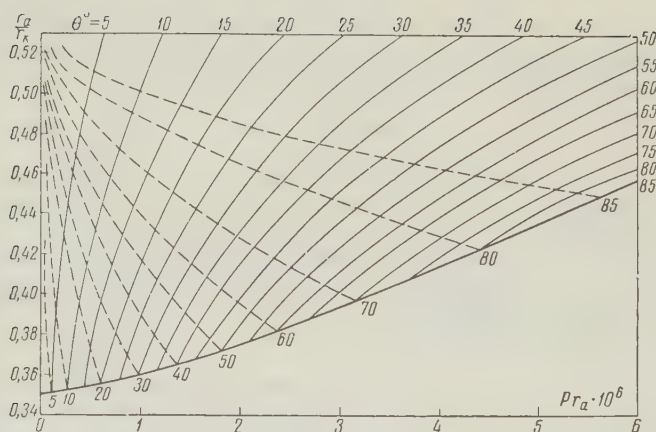


Figure 4

boundary of the region of existence of the constricted stream. The curve joining these points is the lower boundary of the region of values r_a/r_k , Pr_a and θ where it is possible to obtain a constricted stream. The upper boundary of this region is defined by the value $r_a/r_k = 0.53$.

On the basis of Equations (5), (6), (7), and (8), Figure 5 shows the values of y_m and z_m as a function of θ at several values of r_a and $P = 4 \cdot 10^4 \frac{a}{v^{3/2}} \text{ m}$.

The data obtained may be used for an approximate calculation of the basic parameters of an electron gun for given characteristics electron stream. Let, for example, the following be given: current I anode voltage U_a , minimum stream thickness $2y_m$, its width l and the permissible cathode current density. Then we determine the perveance of the flow according to the relation $P = I_a / l U_{a2}^3$, we choose the anode radius r_a and we determine Pr_a . From the curve of Figure 4 we find for a give Pr_a the range of possible values θ in the focusing regions and from Figure 5 we choose the value θ , which corresponds to a prescribed magnitude of y_m and also we determine z_m . Returning again to the curves of Figure 4 we find the sought-for relation r_a/r_k which corresponds to the resulting value of θ . We find $y_0 = r_a \tan \theta$.

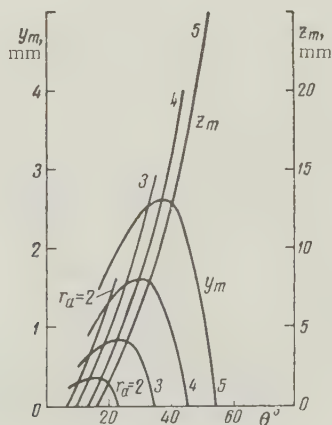


Figure 5

Electron guns having the electrode shape indicated, with an anode aperture transverse cross section of $1 \times 8 \text{ mm}$, for $2\theta = 30^\circ$, $r_a/r_k = 0.36$, $r_a = 2 \text{ mm}$, have in fact yielded

ribbon electron streams in which the current $I = 50$ ma ($U_a = 800-1000$ v) when the current transmission coefficient was 90-95% and the operating conditions were close to the design conditions.

REFERENCES

1. J.R. Pierce, Rectilinear electron flow in beams, J. Appl. Phys., 1940, 11, 8, 548.
2. D.E. Radley, The theory of the Pierce type electron gun, J. Electronics and Control, 1958, 4, 2, 125-148.
3. M.D. Gabovich, Effect of space charge during the propagation of intense beams of charged particles, Uspekhi Fiz. Nauk 56, 2, 215 (1955).
4. H. Rothe, W. Kleen, Hochvakuum Elektronenröhren, Akademische Verlagsgesellschaft M.B.H., Frankfurt a/M, 1955, 1, 161.

Received by the editors February 18, 1960.

SOME SPECIAL FEATURES OF FIELD EMISSION CATHODE OPERATION IN A MICROWAVE FIELD

M.I. Elinson and V.A. Gor'kov

Field emission from a metal or a semiconductor can be obtained not only by the use of an electrostatic field, but also with electrical fields which change with time.*

In 1958 we suggested a system in which the field emission electrons from the point were placed directly in the microwave fields of resonators or of high-power waveguides, and analyzed the possibilities of such systems [1] (see, for example, Figures 1 and 2).

Recently publications have appeared concerning the creation of a number of important types of electron devices based on similar concepts [2].

The goal of this paper is a qualitative theoretical analysis of certain features of the operation of field emission cathodes in microwave electrical fields, which were not considered in Reference 2.

1. It is obvious that the field emission of a cathode operating in a sinusoidal electric field, as a result of the strong nonlinearity of the volt-ampere characteristic, will occur in the form of transitory electron packets.

One can show that $\sim 42\%$ of the charge emitted during a period is concentrated within the phase range 0.09π (16°). During this interval changes in the electric field amount to $\Delta E = \pm 0.5\%$. Numerical integration of the one-dimensional (for simplicity) relativistic equation of motion $m\ddot{x} / \sqrt{1 - (\dot{x}/c)^2} = E_a \sin \omega t$ allows a calculation of the dimensions of the electron packets, their duration and space charge distribution.

In the phase interval $(0.4-0.6)\pi$ at $E_a = 10^7$ v/cm, $d = 0.1$ cm and $\omega = 2\pi \cdot 10^{10}$ sec $^{-1}$, the spread in the time of electron flight in the oscillator is $\Delta t = 0.6\%$, and the velocity spread is $\Delta v = 0.4\%$. The packet extension at the resonator output is $\Delta l = 3$ mm and its duration $\Delta \tau = 10^{-11}$ sec [over the phase range $(0.45-0.55)\pi$ $\Delta l = 1.5$ mm and $\Delta \tau = 5 \cdot 10^{-12}$ sec].

*Therefore use of the term "electrostatic emission" may sometimes lead to misunderstanding.

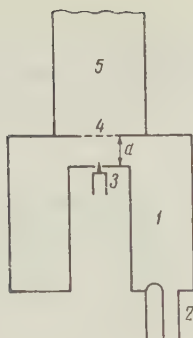


Figure 1

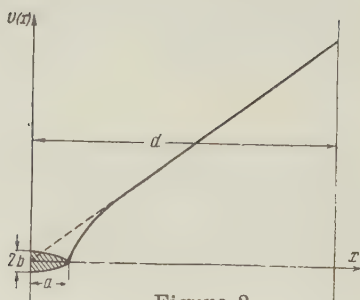


Figure 2

Figure 1. Point in a toroidal microwave resonator: 1-resonator; 2-power input; 3-point with a heating device; 4-fine-mesh grid in the resonator wall; 5-equipotential space in which electron packets are propagated; d is the width of the resonator gap.

Figure 2. Potential distribution in the oscillator gap with the point (solid curve) and without the point (dashed curve). At $a = 10 \text{ v}$ $E_0 \approx 50 \text{ E}$ [3], where E is the uniform field in the resonator without a point. At $E \approx 10^6 \text{ v/cm}$ $E_0 \approx 5 \cdot 10^7 \text{ v/cm}$ and is sufficient to produce high field emission current densities.

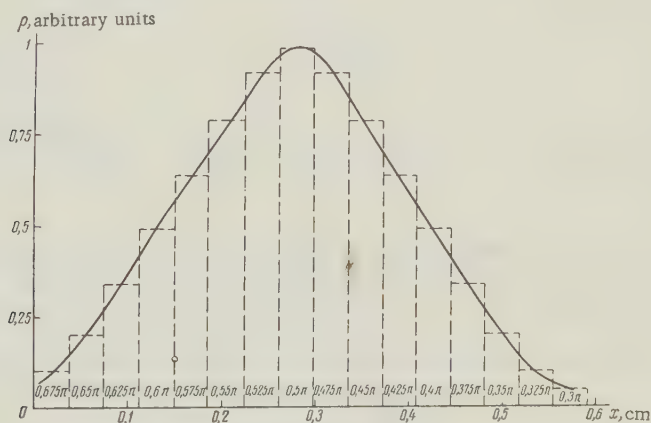


Figure 3. Distribution of space charge density in an electron packet at the resonator output.

Figures 3-5 show the space charge distribution in the electron packets $\rho = \sum_{0.4\pi}^{0.6\pi} \frac{\Delta i_{\phi}}{v_{\phi}}$ at various distances from the resonator. The concentration of electron charge in the packet head is advantageous for radiating microwave systems. An electron beam in the form of packets can also be used in accelerators, klystron-type devices, etc.

The concentration of electrons in the packet varies in the range from 10^{11} to 10^{14} cm^{-3} depending upon the magnitude of the focusing magnetic field, the degree of ionic compensation (if used), and the microwave power.

2. During operation of a field emission cathode in a microwave system, the character of its total bombardment changes substantially.

Assuming the angle of flight of electrons in the resonator to be negligibly small, and assuming that the ionization cross-section is independent of electron coordinates, in the nonrelativistic approximation (for ions), one may solve the problems of the number of ions incident per unit time on the cathode and of their energy. By elementary means one may obtain

$$n_i = \frac{I_{cp} p}{2ekT} S_i d. \quad (1)$$

Here n_i is the number of ions impinging on the cathode per sec; I_{cp} is the time-average electron current; p is the gas pressure; S_i is the ionization cross-section; d is the gap width of the microwave resonator. It is easy to see that n_i is half the number of ions incident on the cathode in the static case which is related to the specific character of ion motion (sum of the forward and oscillatory motion).

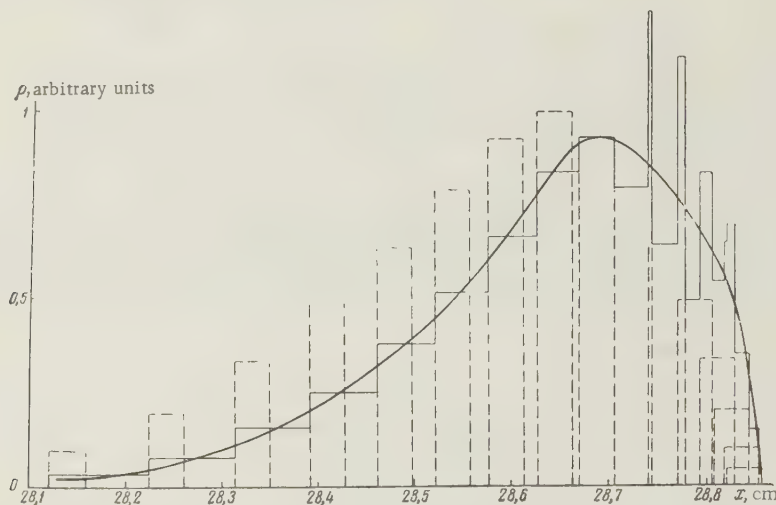


Figure 4. Space charge distribution in the electron packet 30 cm from the resonator.

The ion velocity is

$$\dot{x} = \frac{eE_0}{m_i \omega} (\cos \omega t_0 - \cos \omega t). \quad (2)$$

Here E_0 is the microwave field amplitude; ω is its angular frequency; ωt_0 is the phase of the electron motion; m_i is the ion mass. Taking into account that the maximum number of ions is formed when $\omega t_0 = \pi/2$, we obtain

$$\bar{G}_i = \frac{e^2 E_0^2}{2m_i \omega^2}. \quad (3)$$

It is significant that the energy decreases with an increase of ion mass, since the heavy ions are particularly effective in the process of cathode sputtering, and that the energy also decreases sharply with an increase of field frequency.

A more accurate calculation of the number of incident ions and their energy leads to the following expression:

$$n_i \simeq \frac{p S_i}{ekT} \frac{I_{max}}{2\pi} \left[0.29d + \frac{0.44eE_0}{m_i \omega^2} \right], \quad (4)$$

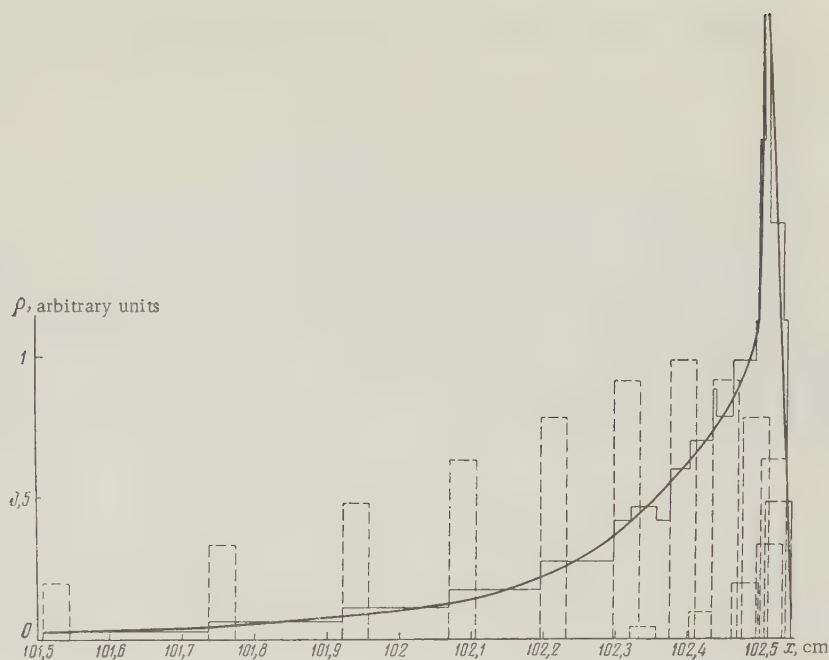


Figure 5. Space charge distribution in the electron packet 100 cm from the resonator.

where I_{\max} is the maximum field emission current,

$$\bar{\mathcal{E}}_i \approx 1.25 \frac{e^2 E_0^2}{2m_i \omega^2}. \quad (5)$$

For $m_i = 100 m_H$, $E_0 = 10^7$ v/cm, $\omega = 7.5\pi \cdot 10^{10} \text{ sec}^{-1}$ ($\lambda = 8$ mm) $\bar{\mathcal{E}}_i \approx 10$ ev; for $m_i = 50 m_H$ (argon) $\bar{\mathcal{E}}_i \approx 20$ ev. It is apparent that the incident ion energy decreases radically in comparison with the static case and this example is close to the threshold of cathode sputtering.

In such devices we virtually exclude the problem of cathode sputtering, and the stability of the field emission increases substantially.

REFERENCE

1. M.I. Elinson and V.A. Gor'lov, Author's. Certificate No. 18430 to 27/V 1958 (Priority, January 16, 1958).
2. R.W. Dyke, IRE Trans., Military Electronics, 1960, 4, 1, 38-45.
3. L.D. Landau and E.M. Lifshits, Electrodynamics of Continuous Media, GITTL (1957), p. 45.

Received by the editors October 19, 1960.

CALCIUM ADSORPTION ON TUNGSTEN

V.I. Makukha

In connection with the wide use of various types of composite cathodes in modern vacuum-tube devices — particularly such cathodes on a base of alkali-earth metal which combine the properties of semiconductor and coated cathodes — great interest has been shown in the investigation of the behavior of films of alkali-earth metals on refractory metals, particularly on tungsten. However, up to the present only the Ba-W system has been investigated [1-8], whereas not a single report has been published on the Ca-W and Sr-W systems, this despite the fact that in many composite cathodes the addition of strontium or calcium compounds has resulted in considerable increase both in the stability of emission from the cathode and in the longevity of the cathode. Hence, investigation of the Ca-W system is of considerable interest in clarifying the role of calcium in the mechanism of increasing emission of composite cathodes.

The present paper discusses the results of an investigation of the Ca-W system by means of an field emission projector [8] with vacuum conditions on the order of $(1-2) \cdot 10^{-9}$ mm Hg. Calcium beryllate in a tantalum boat was used as the calcium source. The calcium was sprayed in small portions on a monocrystal of tungsten from below and by careful heating, the individual portions were dispersed over the entire surface of the point. After dispersion of each portion the image was "frozen" and the current and voltage were measured and the emission pattern was photographed.

The relative work function was determined from the slope of the characteristic [8]

$$\lg \frac{I}{V^2} = f(1/V).$$

1. EXPERIMENTAL RESULTS

Curves for the dependence of the work function φ on the duration of calcium deposition t_s were plotted for several tungsten points. Figure 1 shows one of these curves. For all the emitters the curves of $\varphi(t_s)$ passed through a minimum of 2.4 ± 0.1 ev.

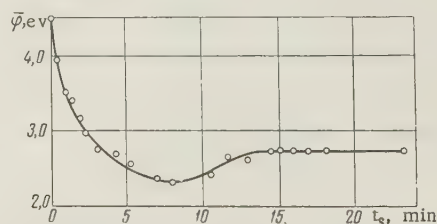


Figure 1. $\bar{\varphi}$ as a function of duration of deposition of Ca on W.

Figure 2 shows a series of patterns with successive spraying (with the point at room temperature) of individual portions of calcium with subsequent dispersion over the entire tungsten surface (at a point temperature not above 450° C).

A thin layer of Ca sprayed on half of the point begins to migrate at 400° C and at 500° C spreads over the entire point in 30 seconds. The calcium avoids faces $\{110\}$ and $\{100\}$. First faces $\{112\}$ and $\{113\}$ are coated (uniting in one bright spot) and the centers of faces $\{111\}$ (Figure 2a) are also coated. In subsequent spraying the Ca spreads over face $\{116\}$, becomes firmly established on $\{111\}$ and evidently forms a narrower thick layer on faces $\{112\}$, for the

latter emit not so strongly and then cease glowing altogether (Figure 2b). Figure 2c corresponds approximately to the optimum coating. The only emitting faces are $\{116\}$ and $\{111\}$ (the latter in the form of bright triangles).

The method of "supplementary spraying" [9] permitted verification of the fact that at the given stage we are dealing with an emission pattern which is characterized by a considerably reduced work function at the optimum coating. The mean work function in this case was 2.4 eV. Spraying of large amounts of Ca results in a work function attributable only to Ca (2.7-2.8 eV) as shown in Figure 2d, wherein it is evident that crystallite formation has begun. For some points the field emission pattern of a thick layer of Ca was slightly different at the same value of work function.

The most typical stages of gradual evaporation of Ca from tungsten at 800° C are shown in Figure 3. After pre-heating for 15-30 sec Ca leaves faces $\{116\}$, appears as bright spots on faces $\{112\}$ and continues to cling firmly to $\{111\}$ (Figure 3a). With further heating faces $\{112\}$ are freed and the triangles of the octahedron decrease in size but continue to emit strongly (Figure 3b). In Figure 3c faces $\{112\}$ are completely free of calcium and faces $\{111\}$ are also almost clear. Only the peripheries of these faces show strong emission. After 15-20 minutes of heating the calcium completely leaves the tungsten and the emission pattern is that of pure tungsten (Figure 3d).

2. EVALUATION OF RESULTS

The experiments revealed that migration of Ca begins at 400° C and evaporation at 700° C. In spraying Ca on tungsten with the latter at room temperature crystallites may form, which disappear at a temperature below 400° C. At none of the stages of spraying or evaporation were traces of calcium observed on faces $\{110\}$. In isolated cases of spraying indistinct annular borders are observed at the faces of the cube.

In evaporation the Ca leaves the faces of the monocrystal in the following order: $\{110\}$, $\{100\}$, $\{112\}$, $\{111\}$ (our discussion is limited to the principal crystallographic axes). Consequently, the sequence of faces in order of decreasing adatom bonds of Ca will be $\{111\}$, $\{112\}$, $\{100\}$, $\{110\}$, that is, the strongest adsorption bonds of calcium occur at faces $\{111\}$.

The use of the various representations [1] leads to the following results.

1. The representation of "close contact" of atoms explains the absence of adsorption of Ca on faces $\{110\}$ and $\{100\}$ but does not explain the presence of adsorption on faces $\{111\}$ and $\{112\}$. Actually, in representing the lattice of Ca atoms on these faces we do not obtain a pattern of close contact.

2. Crystallogeometric consideration (further complicated by the requirement of preserving in the adsorbed film the lattice constant characteristic of the adsorbed substance in the crystal state) cannot serve to explain the adsorption of Ca on faces $\{111\}$ and $\{112\}$. This is easily seen from the data of Table 3 in Reference 1.

3. Proceeding from the approach developed by Drechsler [7] and using his plot for the energy of association at the tungsten crystal facets as a function of the values of the metal radius of the adatom ρ and of the sublayer atom r , we may obtain for Ca the following sequence of faces in order of decreasing association of a single atom: $\{123\}$, $\{112\}$, $\{112\}$, $\{111\}$, $\{100\}$, $\{110\}$ — which also fails to explain the preferential adsorption of Ca at faces $\{111\}$. The adsorption at faces $\{112\}$ may be explained by the fact that the association of a single atom of Ca in the "wells" of this face is greater than at the "valleys" and in the "wells" of other faces of the tungsten monocrystal. Evidently, in the case of the adsorption of Ca Drechsler's single-atom theory may be applied to faces $\{112\}$ with greater success than in the adsorption, for example, of Ba, which does not result in such a "porous" film as does Ca.

Comparison of the results of experiment with the crystallogeometric representations for the cases of adsorption of calcium and barium [1] is given in the table.

The results permit explanation of the increase in emission of a barium film on tungsten (described in Reference 10) upon introducing calcium into it by filling with calcium those faces of the tungsten monocrystal on which the barium is not adsorbed. This phenomenon may also partially explain the role of calcium in composite cathodes on a barium base.

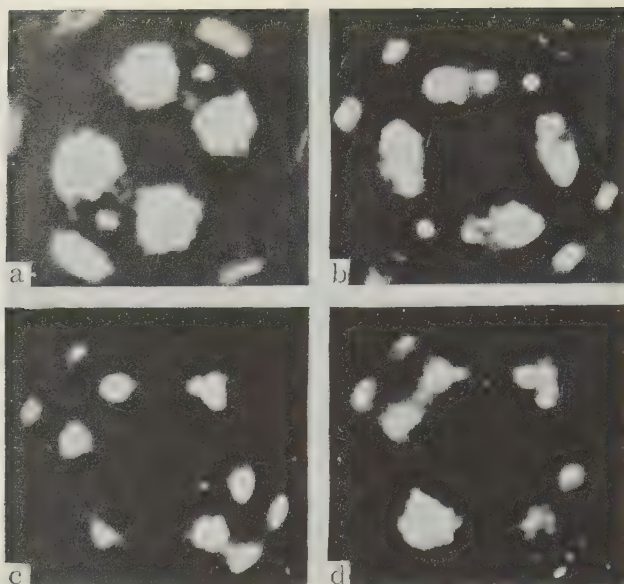


Figure 2. Emission patterns of calcium sprayed on tungsten.
 a - Spraying for 1.5 min, $\varphi = 3.4$ ev; b - spraying for 4 min, $\varphi = 3.1$ ev; c - spraying for 10 min, $\varphi = 2.4$ ev; d - spraying for 25 min, $\varphi = 2.7$ ev.

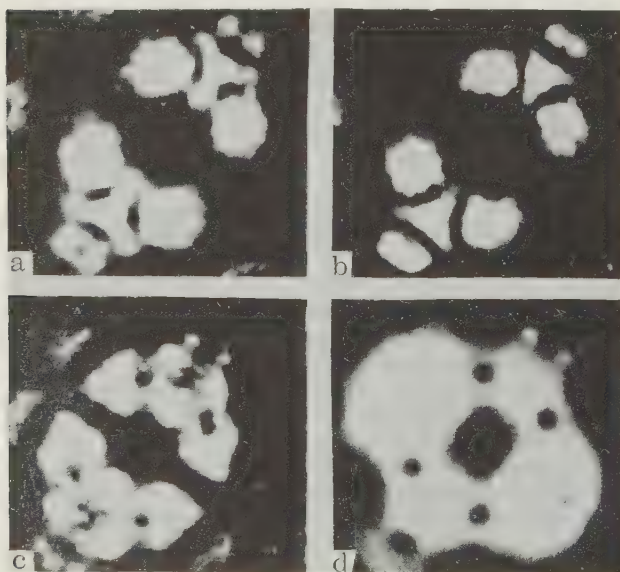


Figure 3. Evaporation of calcium from tungsten (at 800°C).
 a - After 15 sec, $\varphi = 3.1$ ev; b - after 40 sec, $\varphi = 3.4$ ev;
 c - after 5 min, $\varphi = 3.9$ ev; d - after 15 min, $\varphi = 4.5$ ev.

Comparison of Experimental Data with Crystallographic Theories of Adsorption

| Adatoms | Lattice constant, Å | Atom diameter, Å | Adsorption in the form of "coherent" films: (1) in compound theory, (2) in "close contact" theory | Sequence of faces in order of decreasing single-atom bond | Test |
|---------|---------------------|------------------|---|---|---|
| Ca | 5.56 | 3.94 | (1) not possible at faces $\{110\}$ and $\{100\}$, $\{112\}$, $\{111\}$ (2) also not possible at these faces | $\{123\}$, $\{112\}$, $\{111\}$, $\{100\}$, $\{110\}$ | at faces $\{111\}$ and $\{112\}$ |
| Ba | 5.02 | 4.5 | (1) not possible at faces $\{110\}$, $\{100\}$, $\{111\}$ (2) possible at faces $\{100\}$, $\{111\}$ | $\{123\}$, $\{112\}$, $\{100\}$, $\{111\}$, $\{110\}$ | at faces $\{111\}$ and $\{100\}$ isolated cases at faces 123 |

REFERENCES

1. G.N. Shuppe, Electronic Emission of Metal Crystals, Izd. SAGU, Tashkent, 1959.
2. E. Müller, Z. Phys., 1943, 120, 270.
3. S.T. Martin, Phys. Rev., 1939, 56, 1.XI, 947.
4. R. Haefer, Z. Phys., 1940, 116, 9/10, 604.
5. M. Benjamin, R.O. Jenkins, Proc. Roy. Soc. A., 1942, 180, 981, 225.
6. J.A. Becker, Bell System Techn. J., 1951, 30, 4, 907.
7. M. Drechsler, J. Elektrochem., 1954, 58, 5, 327.
8. V.N. Shrednik, Izv. AN SSSR, Ser. Fiz., 1958, 22, 5, 594.
9. V.N. Shrednik, Fizika tverdogo tela, 1959, 1, 7, 1134.
10. I. Brodie, R.O. Jenkins, Proc. Phys. Soc. B. 1956, 69, p. 12, 444B, 1343.

Moscow Physical Engineering Institute
Chair of Electronic and Ionic Devices

Submitted to the editors August 29, 1960

ADSORPTION OF STRONTIUM ON TUNGSTEN

V.I. Makukha

The method described in Reference 1 was used for an investigation of strontium on a tungsten monocrystal.

For several tungsten points curves were plotted for the dependence of the work function $\bar{\varphi}$ on the duration of strontium spraying. One of these curves is shown in Figure 1. The curve $\varphi(t_s)$ passes through a minimum at ~ 2.2 ev.

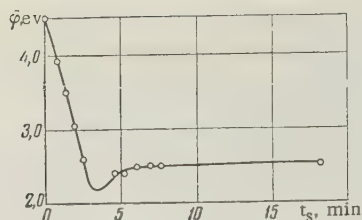


Figure 1. Dependence of $\bar{\varphi}$ on the duration of strontium spraying on tungsten.

Figure 2 shows a series of patterns with sequential spraying of individual portion of strontium with subsequent spreading over the entire surface of the point.

A thin layer of Sr sprayed on half the point begins to migrate at 400–500° C. First faces {233} are coated and then the periphery of the cube. In subsequent spraying the strontium is firmly established on {111} (Figure 2a). New portions of Sr lead to the pattern shown in Figure 2b, wherein emitting faces {111} and {233} create an emission pattern in the form of triple lobes. With further spraying these lobes become bright right triangles at {111}. The photograph in Figure 2c corresponds to the approximately optimum coating. Faces {111} and {233} emit (the lobes are rotated, as it were, through 180° about an axis passing through the center of face {111} and the periphery of faces {112} emit in the form of bright rings. The method of supplementary spraying [2] confirmed the presence of an optimum layer in the given case. The mean work function in this case was ~ 2.2 eV.

After further spraying, the slope of the characteristic of $\lg \frac{I}{V^2} = f(1/V)$ for a thick layer of strontium (see Figure 2d) yielded a work function of $\bar{\varphi} = 2.5$ –2.6 eV.

The most typical stages of evaporation of Sr from thick coatings at 800° C are shown in Figure 3. After 10–15 sec of heating the Sr is firmly established on faces {233}, bright spots appear on {113}, the centers of the octahedron faces become dark (Figure 3a). Further heating results in the formation of bright triangles on {111} (Figure 3b), which are extremely stable in subsequent stages of evaporation. Subsequent evaporation frees the adatoms from faces {111} (Figure 3c). The image is characteristic for the stage preceding total evaporation of the strontium. After 20 min of heating at 900° C the Sr completely leaves the tungsten monocrystal and the point image acquires the appearance characteristic of pure tungsten (Figure 3c).

EVALUATION OF RESULTS

Migration of strontium begins at 450–500° C, evaporation at 750–800° C. During the spraying of strontium on a tungsten point at room temperature crystallites are formed which disappear at $\sim 400^\circ$ C. Traces of Sr on faces {110} were not observed at any stage of spraying or evaporation. Around the faces of the cube the strontium is arranged in the form of a bright "cross." There were observed cases of extremely short-lived coatings of faces {112} with strontium. The emission patterns at different stages of strontium evaporation are similar to the images at certain stages of spraying. In order of decreasing Sr adatom bonds the tungsten faces may be listed in the following sequence: {111}, {113}, {112}, {100}, {110}, that is, the firmest adsorption bonds of strontium occur at faces {111}.

The use of certain representations [3] leads to the following results.

1. Proceeding from Drechsler's point of view [4] and using the plots prepared by him, we may obtain for Sr the following sequence of faces in order of decreasing single-atom bonds: {123}, {112}, {100}, {111}, {110}, which does not explain the preferential adsorption of Sr on faces {111}.

2. Crystallogometric considerations (further complicated by the requirement of preserving in the adsorbed film the lattice constant characteristic of the absorbed substance in

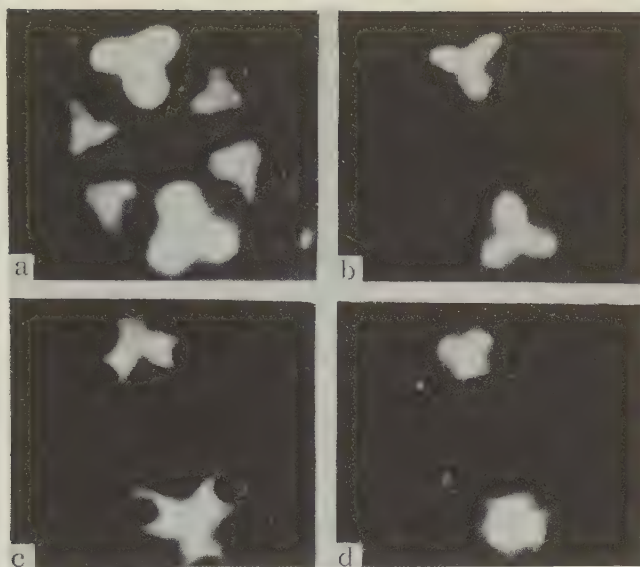


Figure 2. Emission patterns of strontium sprayed on tungsten

a - After 1.5 min, $\bar{\varphi} = 3.5$ ev; b - after 2 min, $\bar{\varphi} = 3$ ev;
c - after 3.5 min, $\bar{\varphi} = 2.2$ ev; d - after 7.5 min, $\bar{\varphi} = 2.5$ ev.

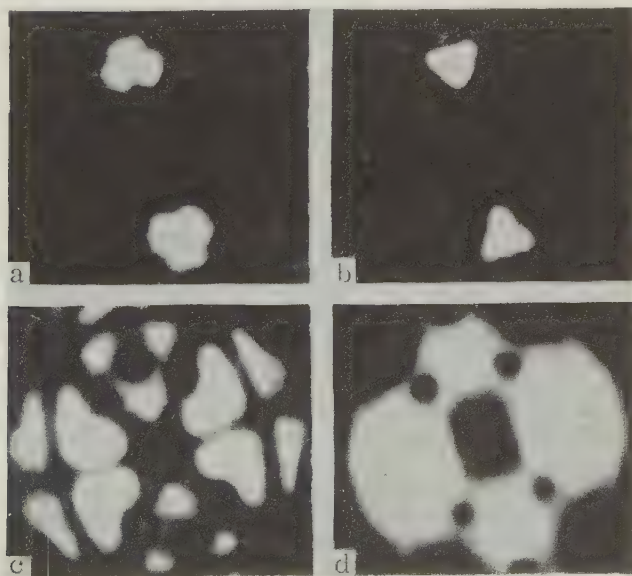


Figure 3. Evaporation of strontium from tungsten (at 800°C).

a - After 10 sec, $\bar{\varphi} = 2.5$ ev; b - after 40 sec, $\bar{\varphi} = 2.6$ ev;
c - after 5 min, $\bar{\varphi} = 4.2$ ev; d - after 25 min, $\bar{\varphi} = 4.5$ ev.

Comparison of Test Data with Crystallographic Theories of Adsorption

| Adatoms | Lattice constant, Å | Atom diameter, Å | Adsorption in the form of "coherent" films: (1) in compound theory, (2) in "close contact" theory | Sequence of faces in order of decreasing single-atom bond | Test |
|---------|---------------------|------------------|--|---|---|
| Sr | 6.075 | 4.3 | (1) not possible at faces {110}, {112}, {100}, {111} (2) possible almost solidly at faces {100} and {111} | {123}, {112}, {100}, {111}, {110} | at face {111} less firmly at faces {223}, {113}, {233} |

the crystal state) cannot serve to explain the adsorption of Sr on faces {111} even in explaining the absence of the latter on faces {100}, {112}, {110}. This is easily seen from the data in table 3 in Reference 3.

3. The representation of "close contact" of atoms explains the presence of adsorption on faces {111} and the absence of the same on such fundamental crystallographic directions as {112} and {110}, but does not explain the absence of adsorption at the faces of the cube. Actually, in depicting the atomic lattice of strontium at faces {111} we obtain their almost "solid" contact (slightly less so than for Ba). Such close contact is obtained also at faces 110 although Sr is not adsorbed at the latter.

The results of the experiment and the application of crystallographic representations for the adsorption of Sr are given in the table (similar data for Ba and Ca are given in Reference 1).

REFERENCES

1. V.I. Makukha, Radiotekhnika i elektronika, 1961, 6, 2, 339.
2. V.I. Shrednik, Fizika tverdogo tela, 1959, 1, 7, 1134.
3. G.N. Shuppe, Electronic Emission of Metal Crystals, Izd. SAGU, Tashkent, 1959.
4. M. Drechsler, Z. Elektrochem, 1954, 58, 5, 327.

Moscow Physical Engineering Institute
Chair of Electronic and Ionic Devices

Submitted to the editors October 3, 1960

DETERMINING THE FREE BARIUM CONTENT IN AN OXIDE CATHODE BY A RADIOACTIVE INDICATOR METHOD

V. P. Vasil'yev and V. P. Pomazkov

Of great interest in the physics of oxide-coated cathodes is the determination of the free barium content in the oxide layer. The methods developed to date have been based on the reaction of water vapor with free barium and the measurement of the pressure of the liberated hydrogen (the Berdennikov method) or the same reaction with nitrogen and other materials [1]. One of the recent experiments performed according to the Berdennikov method [2] had a sensitivity of $1 \cdot 10^{-8} \text{g}$.

The method proposed by us is based on the reaction of free barium with one of the halogens — iodine. The result of the reaction is a somewhat stable compound of BaI_2 . The radioactive isotope of iodine I^{131} was used in the form of an aqueous solution without a carrier which permitted a considerable increase in the sensitivity of the method.

The following procedure was used in purifying the iodine. First, utilizing the excellent solubility of iodine in benzene, several drops of aqueous solution of iodine were mixed with benzene and the latter was precipitated from the water. Here we determined the amount of stable iodine (the carrier) required to create the necessary vapor pressure in test vacuum tube. The solution of iodine in benzene was placed in a glass tube with 4-5 outlets and the iodine was subjected to vacuum distillation. The end of the tube in which the iodine finally condensed had a sealed capillary with a striker. Before the iodine was completely distilled the entire tube was carefully degassed. Then the tube with the radioactive iodine was welded onto the test vacuum tube.

The vacuum tube was a diode with an end cathode of diameter 5 mm and a plane anode. The material for the core was purest nickel (99.99%) annealed in vacuum. The face of the cathode was coated with a double carbonate: 55% BaCO_3 , 45% SrCO_3 on an organic binder. The thickness of the coating was approximately 100-120 microns. The cathode was activated under a vacuum of not more than $5 \cdot 10^{-5} \text{ mm Hg}$ (that is, under ordinary plant conditions) in two stages. In the first stage the temperature reached 850°C and in the second stage 1050°C with testing of the emission current. In static operation the current reached 300-500 ma/cm^2 . The cathode was then cooled and the tube was sealed off at a vacuum of 10^{-7} mm Hg .

Immediately after sealing off the tube the capillary was broken and the iodine vapor penetrated into the tube. The tube was placed in an oven at a temperature of $100\text{--}120^\circ \text{C}$ for several hours, after which the oven temperature was raised to 200°C and the end of the ampule was cooled. As a result all the uncombined iodine collected at one point in the device. After this the tube was opened and the radioactivity of the cathode and anode was determined. If the distribution of radioactivity in the depth of the oxide layer was to be determined, the cathode was coated with paraffin and slices of 10-micron thickness were cut with a special microtome.

The high specific activity of I^{131} permits reliable measurement of its quantity up to 10^{-10}g . From the amount of iodine in the oxide coating it is easy to determine the amount of free barium. The average measured amount of free barium on the cathodes was $5 \cdot 10^{-7} \text{g}$, which corresponds to a concentration of approximately 0.05% relative to the weight of the entire layer. Moreover, free barium was detected on the anode surface facing the cathode and was measured at $3 \cdot 10^{-8} \text{g}$.

In order to determine iodine absorption in the oxide layer control tests were performed. For this purpose a tube was constructed with two oxide cathodes and two anodes. One of the cathodes was activated to the point of stable emission (as described above), after which oxygen was injected into the tube at a pressure of several atmospheres. After several hours the

tube was again evacuated and the second cathode was activated, after which the usual measurements with radioactive iodine were performed. The amount of free barium on the uncontaminated cathode was close to that given in the data above, while on the contaminated cathode it did not exceed $1 \cdot 10^{-9}$ g and at the contaminated anode $5 \cdot 10^{-10}$ g.

The results show that absorption of iodine in the oxide layer is two orders of magnitude less than the amount of iodine associated with free barium in the layer. The data on the distribution of free barium in the oxide layer indicate an increase in its concentration from surface to core. The amount of free barium at the anodes is apparently smaller due to the fact that the vacuum conditions permit part of the free barium which evaporates from the cathode to combine with residual gases. As a result of the use of tagged atoms in determining the concentration of free barium in an oxide-coated cathode it may be stated that this method is at least twice as sensitive as other existing methods.

REFERENCES

1. G. German, S. Vagener. The Oxide-Coated Cathode, GTI, 1949.
2. P.V. Kovtunencko, B.M. Tsarev, Radiotekhnika i elektronika, 1959, 4, 5, 866.

Tashkent State University
Chair of Electrophysics

Submitted to the editors July 6, 1960.

RADIO ENGINEERING AND ELECTRONIC PHYSICS

Institute of Radio Engineering and Electronic Physics,
Academy of Sciences of the USSR

EDITORIAL BOARD

Editor-in-Chief: V.A. Kotel'nikov

Associate Editors: D.V. Zernov, Yu.B. Kobzarev

| | | |
|------------------|------------------|----------------|
| A.I. Berg | L.N. Dobretsov | A.M. Prokhorov |
| B.A. Vvedenskiy | A.N. Kazantsev | S.M. Rytov |
| I.S. Gonorovskiy | S.G. Kalashnikov | V.I. Siforov |
| V.L. Granovskiy | P.L. Kapitsa | Ya.N. Fel'd |
| L.A. Zhekulin | V.V. Migulin | S.E. Khaykin |
| N.D. Devyatkov | A.L. Mikaelyan | B.M. Tsarev |
| | A.A. Pistol'kors | |

Scientific Secretary of Editorial Board: G.A. Bernashevskiy

The English Edition of Radio Engineering and Electronic Physics is mailed to subscribers within 18 weeks after the publication of the original Russian issue.

Russian electronic journals published by the
American Institute of Electrical Engineers
Translated by Royer and Roger, Inc.

| | Subscription rates | | | |
|---|--------------------|----|-----------|----|
| | Individuals | | Libraries | |
| | \$ | £ | \$ | £ |
| <i>Radio Engineering and Electronic Physics</i> | 28.50 | 10 | 57.00 | 20 |
| <i>Radio Engineering</i> | 14.25 | 5 | 28.50 | 10 |
| <i>Telecommunications</i> | 14.25 | 5 | 28.50 | 10 |

Royer and Roger translates and produces
the following Russian scientific journals:

| | |
|---------------------------------|-------------------------------|
| <i>Biophysics</i> | <i>Problems of Oncology</i> |
| <i>Entomological Review</i> | <i>Radio Engineering</i> |
| <i>Geochemistry</i> | <i>Radio Engineering and</i> |
| <i>Geodesy and Cartography</i> | <i>Electronic Physics</i> |
| <i>Izvestiya, Academy of</i> | <i>Refractories</i> |
| <i>Sciences of the USSR,</i> | <i>Sechenov Physiological</i> |
| <i>Geologic Series</i> | <i>Journal of the USSR</i> |
| <i>Pavlov Journal of Higher</i> | <i>Soil Science</i> |
| <i>Nervous Activity</i> | <i>Telecommunications</i> |

Comments and inquiries regarding *Radio Engineering and Electronic Physics* and other translation journals should be sent to:

International Division
Royer and Roger, Inc.

1000 Vermont Avenue, N.W.
Washington 5, D.C.

41 East 28th Street
New York 16, New York

C O N T E N T S

| | Page |
|--|------|
| Yu.S. Lezin: Noise Storage in Delayed Feedback Devices | 159 |
| P. Ya. Nudel'man: Linear Correction of Transients in Systems with Finite Passband | 165 |
| Lin Hai-Ts'uan: Calculation of Channel Capacity in Two-Path Communication | 168 |
| Lin Hai-Ts'uan: Capacity of One-Path and Two-Path Communications Channels with Polarization Fading | 172 |
| A. V. Shabel'nikov: The Perturbation Method and its Relation to the Exact Diffraction Method in Solving Problems of Incoherent Scattering | 176 |
| N. G. Ponomarev: Graphical Method of Designing Aplanatic Antennas | 185 |
| O. N. Tereshin and A. Ye. Sokolov: Suppression of Currents Excited in Metal Shields by Diffraction Antennas of Finite Dimensions | 191 |
| B. Z. Katsenelenbaum, and Z. A. Malina: Transition Calculations for a Symmetrical Magnetic Wave in a Circular Waveguide | 198 |
| A. L. Fel'dshteyn: Synthesis of Stepped Directional Couplers | 204 |
| Yu. N. Kazantsev: Attenuation Measurements in Waveguides | 210 |
| V. I. Muromtsev and A. K. Piskunov: Some Problems of Microwave Channel Matching in Radio Spectroscopes | 218 |
| G. N. Rapoport: Large-Signal Operation of Backward-Wave Amplifier | 227 |
| L. V. Kasatkin: Amplification of Space-Charge Waves in the Passage of Electron Beams within Media with Inductive Susceptance | 232 |
| A. L. Igritskiy: Calculation of a Magnetic Periodic Focusing Device for TWT | 239 |
| G. D. Lobov, and V. I. Yeremeyev: Some Effects Accompanying Detection in a Gaseous Discharge | 248 |
| M. I. Elinson, G. V. Stepanov, and V. I. Pokalyakin: Emission of Hot Electrons from p-n Junctions in SiC Crystals. | 252 |
| G. N. Fursei : Field Emission from Tungsten Single Crystals in the Prebreakdown Current Region | 257 |
| G. Ya. Myakishev: On the Theory of the Static Characteristics of a Planar Diode without an Incandescent Cathode taking into account Thermal Electron Velocity Distribution | 262 |
| Yu. R. Nosov: Transition Processes in Semiconductor Diodes Using Narrow Forward Current Pulses | 271 |
| V. V. Grigor'yants, and M. Ye. Zhabotinskiy: Molecular Frequency Standard with Subtraction of Reference Oscillator Error | 279 |
| BRIEF COMMUNICATIONS | |
| L. K. Mikhaylovskiy: Wideband Amplification with Frequency Conversion . . | 286 |
| A. I. Yel'kind: Suppression of Harmonics in Coaxial Resonators | 288 |
| L. E. Bakhrakh and A. D. Sulimin: Calculation of Ribbon Stream Electron Guns | 290 |
| M. I. Elinson and V. A. Gor'kov: Some Special Features of Field Emission Cathode Operation in a Microwave Field | 294 |
| V. I. Makukha: Calcium Adsorption on Tungsten | 298 |
| V. I. Makukha: Adsorption of Strontium on Tungsten | 301 |
| V. P. Vasil'yev, and V. P. Pomazkov: Determining the Free Barium Content in an Oxide Cathode by a Radioactive Indicator Method | 305 |